Statistical Inference in Dynamic Treatment Regimes

Eric B. Laber, Daniel J. Lizotte, Min Qian,

William E. Pelham and Susan A. Murphy

November 10, 2011

Statistical Inference in Dynamic Treatment Regimes

Eric B. Laber, Daniel J. Lizotte, Min Qian,

William E. Pelham, and Susan A. Murphy[1]

Abstract

Dynamic treatment regimes, also known as *treatment policies*, are increasingly being used to operationalize sequential clinical decision making associated with patient care. Common approaches to constructing a dynamic treatment regime from data, such as Q-learning, employ non-smooth functionals of the data. Therefore, simple inferential tasks such as constructing a confidence interval for the parameters in the Q-function are complicated by nonregular asymptotics under certain commonly-encountered generative models. Methods that ignore this nonregularity can suffer from poor performance in small samples. We construct confidence intervals for the parameters in the

[1]Eric B. Laber is in the Department of Statistics at North Carolina State University, 2311 Stinson Dr., Raleigh, NC, 27695 (E-mail *laber@stat.ncsu.edu*). He acknowledges support from NIH grant P01 CA142538. Daniel J. Lizotte is in the Department of Computer Science at the University of Waterloo, Ontario, N2LG31. Min Qian is in the Department of Biostatistics at Columbia University, 722 West 168th Street, New York, NY 10032. William E. Pelham is Professor of Psychology and Directory of Center for Children and Families at State University of New York at Stony Brook, Buffalo, NY 14260. Susan A. Murphy is in the Departments of Statistics and Psychiatry at the University of Michigan, Ann Arbor, MI, 48109. She acknowledges support from NIMH grant R01-MH-080015 and NIDA grant P50-DA-010075.

Q-function by first constructing smooth, data-dependent, upper and lower bounds on these parameters and then applying the bootstrap. The confidence interval is adaptive in that although it is conservative for nonregular generative models, it achieves asymptotically exact coverage elsewhere. The small sample performance of the method is evaluated on a series of examples and compares favorably to previously published competitors. Finally, we illustrate the method using data from the Adaptive Interventions for Children with ADHD study (Pelham and Fabiano 2008).

1 Introduction

Dynamic treatment regimes, also known as *treatment policies*, are increasingly being used to explore how to inform sequential clinical decision making using data. Clinical scientists, wanting to develop principled, evidence-based rules for tailoring treatment, have conducted Sequential, Multiple Assignment, Randomized Trials (SMART; Lavori and Dawson 2003; Murphy 2005; Murphy et al. 2007) in order to evaluate and compare different long-term dynamic treatment regimes. In this work, we develop confidence interval methodology that can be used to address the following types of scientific questions arising in the development of dynamic treatment regimes: "Is there sufficient evidence to conclude that a particular treatment is best compared to other treatments when these treatments are considered in the context of a dynamic treatment regime?," "Is a particular patient variable useful in tailoring treatment, and for which configuration of the patient variables is there sufficient evidence to conclude that there exists a unique best treatment option?"

This work is motivated by our involvement in the Adaptive Interventions for Children with Attention Deficit Hyperactivity Disorder (ADHD) study (Center for Children and Families, SUNY at Buffalo, William E. Pelham PI, IES Grant R324B060045; see also Nahum-Shani et al. 2010a). ADHD affects an estimated 5%-10% of school aged children, and is characterized by inattention, hyperactivity, and impulsivity (Pliszka 2007). In the years

preceding the study, clinicians debated the comparative effectiveness of behavioral modification therapy versus medication as treatment options for ADHD as well as the best sequencing of these treatments (Pliszka 2007; Pelham and Fabiano 2008). As a consequence, a SMART trial was conducted with the general aim of estimating the dynamic treatment regime that achieves the greatest reduction in ADHD symptoms among school age children. This SMART study is composed of two stages. In the first stage, children were randomized with equal probability into one of two treatment groups (low-dose behavioral modification therapy, low-dose medication). After a burn-in period of eight weeks, children were evaluated monthly and at each evaluation deemed either a responder or non-responder. (The operationalized definition of nonresponse is given in Nahum-Shani et al. (2010a)). Non-responders were immediately re-randomized to either (i) augmentation of treatment, so that the child was provided both medication *and* behavioral modification therapy, or (ii) intensification of treatment, so that the child was provided an increased dosage of their current (stage one) treatment. Responders were not re-randomized and were provided their current treatment at the current dosage level.

Data collected in a SMART trial like the ADHD study can be used to estimate an optimal dynamic treatment regime. This estimation typically uses an extension of regression to multistage decision making problems. The extension we consider in this paper is the Q-learning algorithm (Watkins and Dayan 1992; Murphy 2005). A variety of other extensions exist in the statistical literature (Murphy 2003; Robins 2004; Blatt et al. 2004; Moodie et al. 2007; Henderson et al. 2009; Zhao et al. 2009). However *all of these extensions* suffer from the same problem of nonregularity that we focus on in this paper (Robins 2004; Moodie and Richardson 2007; Henderson et al. 2009; Chakraborty et al. 2009; Moodie et al. 2010).

In this paper, we provide a method for constructing confidence intervals for parameters arising in the Q-Learning algorithm. The primary challenge to this task is that the estimators are non-smooth functionals of the data—in particular, the formula for the estimators involves

the use of the max operator, which is non-differentiable. Robins (2004) notes two problems resulting from the non-differentiability of the max operator. First, while the estimators of the regression coefficients are consistent, their limiting distributions can have nonzero mean; that is, there is estimation bias on the order of $1/\sqrt{n}$ for some generative models. Second, the regression coefficient estimators are nonregular (Bickel 1993; Tsiatis 2006). That is, their limiting distributions changes abruptly as one smoothly varies the underlying generative model. As a practical consequence, common approaches based on the bootstrap and Taylor series arguments provide inconsistent interval estimators and can behave poorly in small samples (Andrews and Ploberger 1994; Andrews 2001, 2002; Leeb and Pötscher 2005).

The adaptive confidence interval proposed here is based on smooth, data-dependent, upper and lower bounds on the estimators involved in the regression models used by Q-learning. Confidence intervals are formed by bootstrapping these bounds. The proposed confidence interval is adaptive in that although it is conservative for nonregular generative models, it achieves asymptotically exact coverage elsewhere.

Many authors have focused on reducing the bias of order $1/\sqrt{n}$ discussed above (recall that under some generative models, the estimators of the regression coefficients can have limiting distributions with nonzero means, thus bias). The methods of Moodie and Richardson (2007), Chakraborty et al. (2009) and Song et al. (2010) reduce the estimation bias via the use of thresholding. As is well-known, the use of thresholding (or penalization that induces variable selection) leads to nonregular estimation (see Leeb and Pötscher 2003, 2005 and references therein; Chatterjee and Lahiri 2011). Moodie and Richardson propose the use of a hard threshold whereas Chakraborty et al. propose a soft-thresholding method that is motivated by an empirical Bayes argument. Song et al. generalize the soft-thresholding method by use of a lasso-like penalization. In all cases, linear combinations of some parameters may be set to zero as a result of the thresholding. In Chakraborty et al. confidence intervals are constructed by use of the bootstrap whereas in Song et al., confidence intervals are produced

via Taylor series arguments. Both methods work well in the simulations provided. However, the standard bootstrap is inconsistent in nonregular settings (Shao 1994; Beran 1997) and confidence intervals based on Taylor series arguments do not capture the variation due to variable selection/thresholding (Pötscher 1991; Leeb and Pötscher 2005). Neither the work of Song et al., nor Chakraborty et al. need be consistent under a local alternatives framework. In contrast, this work provides regular (consistent under local alternatives) confidence sets.

Instead of focusing on bias reduction of the estimator we focus directly on the construction of high quality confidence intervals. We do this for several reasons: first, it is known that in settings in which there is no unbiased estimator, attempts to eliminate the bias at some parameter values must lead to large mean square error at other parameter values (Doss and Sethuraman, 1989; Liu and Brown 1993; Chen 2004; Hirano and Porter 2009). Simulations provided in the supplementary material (Section 3 of the Supplement) provide examples of this excessive mean square error. Second, interval estimators (such as confidence intervals) that obtain the desired level of confidence can be used to conduct inference about the parameters even when there is bias of the order $1/\sqrt{n}$. Alternate approaches that focus directly on confidence intervals include two proposals by Robins (2004), the first of which is a projection confidence interval. Robins' second proposal and a natural method that we call "plug-in pretesting estimation" share some conceptual similarities with the adaptive confidence interval proposed here. See Section 4.2 for discussion.

Section 2 considers the simplest possible setting, in which there are two stages of treatment and two treatments available at each stage. Here the adaptive confidence interval (ACI) is introduced and asymptotic properties are provided. Section 3 generalizes the problem and the ACI to the class of problems with two stages of treatment and an arbitrary number of treatments at each stage. In Section 4, we provide an empirical comparison of the ACI with the bootstrap and the use of thresholding as represented in Chakraborty et al. (2009) on a

number of test cases. The ACI compares favorably. Section 5 contains an application of the ACI to the analysis of the ADHD study and a discussion of future work. An extension of the ACI to an arbitrary number of stages of treatment, and an arbitrary number of treatments at each stage, is given in the supplementary material.

2 Two stages of binary treatment

In this section, we develop the adaptive confidence interval for the parameters in the Q-function when there are two stages of treatment and two treatments are available at each stage. We use uppercase letters such as X and A to denote random variables, and lowercase letters such as x and a to denote instances of these random variables. The data consist of n trajectories drawn $i.i.d.$ from some fixed and unknown distribution P. Each trajectory $(X_1, A_1, Y_1, X_2, A_2, Y_2)$ is a sequence of random variables collected at two stages $t = 1, 2$; $X_t \in \mathbb{R}^{p_t}$ denotes patient measurements collected prior to the tth assignment of treatment, $A_t \in \{1, 2\}$, denotes the binary treatment (also called an *action*) assigned at stage t and $Y_t \in \mathbb{R}$ is a measure of patient response following the assignment of treatment at stage t. We assume that Y_t has been coded so that a higher value corresponds to a better clinical outcome. Let $H_t = \{X_1, A_1, \ldots, X_t\}$ be the patient history, e.g., the information available to the decision maker *before* the assignment of the tth treatment A_t. Furthermore, we assume that the treatments, A_t, are randomly assigned to patients at each stage with known probabilities possibly depending on patient history.

We wish to use data like the above to inform the construction of a Dynamic Treatment Regime (DTR). A DTR is sequence of decision rules, one for each stage of treatment, that takes as input the patient history and gives as output a recommended treatment. More formally, a DTR $\pi = (\pi_1, \pi_2)$ is an ordered pair of functions π_t so that $\pi_t : \mathcal{H}_t \mapsto \{1, 2\}$ where $\mathcal{H}_t \subseteq \mathbb{R}^{d_t}$ is the domain of H_t. Let $\mathbb{E}^\pi$ denote the joint expectation over H_t, A_t, Y_t

for $t = 1, 2$ under the restriction that $A_t = \pi(H_t)$. The objective is to learn a DTR π which comes close to maximizing the expected clinical outcome $\mathbb{E}^{\pi}(Y_1 + Y_2)$. One way to estimate an optimal DTR is using the Q-learning algorithm (Watkins 1989), which can be conceptualized as an extension of regression to multistage decision making. More precisely, Q-learning is a form of approximate dynamic programming, where the conditional mean responses are estimated from the data since they cannot be computed explicitly. We now describe the Q-learning algorithm with function approximation as in Murphy (2005). To start, define

$$\begin{aligned} Q_2(h_2, a_2) &\triangleq \mathbb{E}\left(Y_2 \middle| H_2 = h_2, A_2 = a_2\right) \qquad (1) \\ Q_1(h_1, a_1) &\triangleq \mathbb{E}\left(Y_1 + \max_{a_2 \in \{1,2\}} Q_2(H_2, a_2) \middle| H_1 = h_1, A_1 = a_1\right); \qquad (2) \end{aligned}$$

the functions $Q_t(h_t, a_t)$, $t = 1, 2$ are known as Q-functions. At each stage of treatment t the Q-function reflects the quality (hence the letter "Q") of the treatment a_t given the patient history h_t. If the conditional expectations in the preceding display were known, then dynamic programming provides an optimal DTR given by $\pi^{dp}(h_t) \triangleq \arg\max_{a_t \in \{1,2\}} Q_t(h_t, a_t)$. In most practical settings these mean functions must be approximated from data. In this paper we consider linear approximations to the conditional mean function. Specifically, we employ a working model of the form

$$Q_t(h_t, a_t; \beta_t) = \beta_{t,0}^{\intercal} h_{t,0} + \beta_{t,1}^{\intercal} h_{t,1} 1_{a_t = 1}, \qquad (3)$$

where $h_{t,0}$ and $h_{t,1}$ are vectors of features comprising the patient history. Note that according to the model, if $h_{t,1}^{\intercal}\beta_{t,1} \approx 0$ then both treatments $a_t = 1$ and $a_t = 2$ yield the approximately same response for a patient with history $H_{t,1} = h_{t,1}$. That is, that there is *not* a unique best treatment for a patient with history $H_{t,1} = h_{t,1}$. Conversely, if $|h_{t,1}^{\intercal}\beta_{t,1}| \gg 0$ then exactly one

treatment yields the best expected outcome for a patient with history $H_{t,1} = h_{t,1}$. We use β_t to denote $(\beta_{t,0}^\intercal, \beta_{t,1}^\intercal)^\intercal$. Let $\mathbb{P}_n$ denote the empirical measure. The Q-learning algorithm proceeds as follows:

1. Regress Y_2 on H_2 and A_2 using (3) to obtain

$$\hat{\beta}_2 \triangleq \arg\min_{\beta_2} \mathbb{P}_n \left(Y_2 - Q_2(H_2, A_2; \beta_2)\right)^2,$$

   and subsequently the approximation $Q_2(h_2, a_2; \hat{\beta}_2)$ to the conditional mean $Q_2(h_2, a_2)$.

2. (a) Define the predicted future reward following the optimal policy as:

$$\begin{aligned}\tilde{Y}_1 &\triangleq Y_1 + \max_{a_2 \in \{1,2\}} Q_2(H_2, A_2; \hat{\beta}_2) &(4)\\ &= Y_1 + H_{2,0}^\intercal \hat{\beta}_{2,0} + \left[H_{2,1}^\intercal \hat{\beta}_{2,1}\right]_+, &(5)\end{aligned}$$

   where $[z]_+$ denotes the positive part of z.

   (b) Regress $\tilde{Y}_1$ on H_1 and A_1 using (3) to obtain $\hat{\beta}_1 \triangleq \arg\min_{\beta_1} \mathbb{P}_n(\tilde{Y}_1 - Q_1(H_1, A_1; \beta_1))^2$.

3. Define the estimated optimal DTR as $\hat{\pi} = (\hat{\pi}_1, \hat{\pi}_2)$ so that

$$\hat{\pi}_t(h_t) \triangleq \arg\max_{a_t \in \{1,2\}} Q_t(h_t, a_t; \hat{\beta}_t).$$

Examination of the above procedure make apparent the close connection between Q-learning and dynamic programming. For further elaboration see Watkins and Dayan (1992), Murphy (2005), and Zhao et al. (2009).

The second stage population coefficients, β_2^*, satisfy $\beta_2^* \triangleq \arg\min_{\beta_2} P\left(Y_2 - Q_2(H_2,\ A_2; \beta_2)\right)^2$. Define $\tilde{Y}_1^* \triangleq Y_1 + H_{2,0}^\intercal \beta_{2,0}^* + \left[H_{2,1}^\intercal \beta_{2,1}^*\right]_+$, then the first stage population coefficients β_1^* are

given by

$$\beta_1^* \triangleq \arg\min_{\beta_1} P\left(\tilde{Y}_1^* - Q_1(H_1, A_1; \beta_1)\right)^2.$$

Notice that β_1^* is a non-smooth function of the trajectories due to the presence of the $\left[H_{2,1}^\intercal \beta_{2,1}^*\right]_+$ term in $\tilde{Y}_1^*$; this term, which is due to the maximization of the stage two Q-function, is the source of the nonregularity in the estimation of optimal dynamic treatment regimes (see Robins, 2004; Chakraborty et al. 2009)

The goal of this paper is the development of asymptotically valid confidence intervals for linear combinations $c^\intercal \beta_1^*$ of the first stage coefficients. Note that standard methods are appropriate for construction of confidence intervals for the second stage coefficients. To better understand the nonregularity and thus the challenge in constructing confidence intervals for the first stage coefficients, we provide a useful decomposition of $c^\intercal \sqrt{n}(\hat{\beta}_1 - \beta_1^*)$. Define $B_1 \triangleq (H_{1,0}^\intercal, H_{1,1}^\intercal 1_{A_1=1})^\intercal$ so that instances of $B_1^\intercal$ form the rows of the design matrix in the first stage regression. Let $\hat{\Sigma}_1 \triangleq \mathbb{P}_n B_1 B_1^\intercal$, then examination of the normal equations shows that $\hat{\beta}_1 = \hat{\Sigma}_1^{-1} \mathbb{P}_n B_1 \tilde{Y}_1$. Hence, for any $c \in \mathbb{R}^{\dim(\beta_1^*)}$ it follows that $c^\intercal \sqrt{n}(\hat{\beta}_1 - \beta_1^*) = c^\intercal \hat{\Sigma}_1^{-1} \sqrt{n} \mathbb{P}_n B_1 \left(\tilde{Y}_1 - B_1^\intercal \beta_1^*\right)$, which, using the definition of $\tilde{Y}_1$, can be further decomposed as

$$c^\intercal \mathbb{W}_n + c^\intercal \hat{\Sigma}_1^{-1} \mathbb{P}_n B_1 \mathbb{U}_n, \tag{6}$$

where

$$\begin{aligned}
\mathbb{W}_n &= \hat{\Sigma}_1^{-1} \sqrt{n} \mathbb{P}_n B_1 \left[\left(Y_1 + H_{2,0}^\intercal \beta_{2,0}^* + \left[H_{2,1}^\intercal \beta_{2,1}^*\right]_+ - B_1^\intercal \beta_1^*\right) + H_{2,0}^\intercal \left(\hat{\beta}_{2,0} - \beta_{2,0}^*\right)\right], \\
\mathbb{U}_n &= \sqrt{n}\left(\left[H_{2,1}^\intercal \hat{\beta}_{2,1}\right]_+ - \left[H_{2,1}^\intercal \beta_{2,1}^*\right]_+\right).
\end{aligned}$$

The second term in (6) is non-smooth which can be seen from the definition of $\mathbb{U}_n$. To illustrate the effect of this non-smoothness, fix $H_{2,1} = h_{2,1}$. If $h_{2,1}^\intercal \beta_{2,1}^* > 0$, then $\mathbb{U}_n\big|_{H_{2,1}=h_{2,1}}$ is easily seen to be asymptotically normal with mean zero. On the other hand, if $h_{2,1}^\intercal \beta_{2,1}^* = 0$,

then $\mathbb{U}_n\big|_{H_{2,1}=h_{2,1}} = \Big[h_{2,1}^{\intercal}\sqrt{n}(\hat{\beta}_{2,1} - \beta_{2,1}^{*})\Big]_{+}$ which converges weakly to positive part of a mean zero normal random variable. Thus, the limiting distribution $c^{\intercal}\sqrt{n}(\hat{\beta}_1 - \beta_1^{*})$ depends abruptly on both the true parameter $\beta_{2,1}^{*}$ and the distribution of patient features $H_{2,1}$. In particular, the limiting distribution of $c^{\intercal}\sqrt{n}(\hat{\beta}_1 - \beta_1^{*})$ depends on the frequency of patient features $H_{2,1} = h_{2,1}$ for which there is no treatment effect (i.e. features for which $h_{2,1}^{\intercal}\beta_{2,1}^{*} = 0$). As discussed in the introduction the nonregularity in the limiting distribution complicates the construction of confidence intervals for linear combinations of β_1^{*}.

The ACI is formed by constructing smooth data-dependent bounds on $c^{\intercal}\sqrt{n}(\hat{\beta}_1 - \beta_1^{*})$. Below we construct upper and lower bounds on $c^{\intercal}\sqrt{n}(\hat{\beta}_1 - \beta_1^{*})$ by means of a preliminary hypothesis test that partitions the data into two sets (i) patients for which there appears to be a treatment effect, and (ii) patients where it appears there is no treatment effect. The bounds are formed by bounding the error of the overall approximation due to misclassification of patients in the partitioning step.

The idea of conducting a preliminary hypothesis test prior to forming estimators or confidence intervals is known as preliminary testing or pretesting (see Olshen 1973); indeed estimators formed by thresholding implicitly use a pretest. Pretesting has been used in Econometrics to provide hypothesis tests and confidence intervals in nonregular settings (Andrews 2001, Andrews & Soares 2007; Cheng 2008; Andrews & Guggenberger 2009). In these settings, one can identify a small number of problematic parameter values (usually one value) at which nonregularity occurs. A pretest is constructed with the null hypothesis that the parameter takes this problematic value. If the pretest rejects, a standard critical value is used to form the confidence interval; if the pretest accepts, the maximal critical value over all possible local alternatives is used to form the confidence interval. In this paper the situation is somewhat different, since nonregularity occurs for any combination of the distribution of the $H_{2,1}$ and $\beta_{2,1}^{*}$ for which $P[H_{2,1}^{T}\beta_{2,1}^{*} = 0] > 0$. Thus we take a different tactic from the simple pretest approach. We conduct a pretest for each *individual* in the

data set as follows. Define $\mathbb{V}_n \triangleq \sqrt{n}(\hat{\beta}_{2,1} - \beta^*_{2,1})$, and $\hat{\Sigma}_{21,21}$ to be the plug-in estimator of the asymptotic covariance matrix of $\mathbb{V}_n$. Each pretest is based on $\hat{T}(h_{2,1}) \triangleq \frac{n(h_{2,1}^\intercal \hat{\beta}_{2,1})^2}{h_{2,1}^\intercal \hat{\Sigma}_{21,21} h_{2,1}}$; note that $\hat{T}(h_{2,1})$ corresponds to the usual test statistic when testing the null hypothesis, $h_{2,1}^\intercal \beta^*_{2,1} = 0$.

The upper bound on $c^\intercal \sqrt{n}(\hat{\beta}_1 - \beta^*_1)$ is given by

$$\begin{aligned}\mathcal{U}(c) \triangleq c^\intercal \mathbb{W}_n + c^\intercal \hat{\Sigma}_1^{-1} \mathbb{P}_n B_1 \mathbb{U}_n 1_{\hat{T}(H_{2,1}) > \lambda_n} \\ + \sup_{\gamma \in \mathbb{R}^{\dim(\beta^*_{2,1})}} c^\intercal \hat{\Sigma}_1^{-1} \mathbb{P}_n B_1 \left(\left[H_{2,1}^\intercal (\mathbb{V}_n + \gamma) \right]_+ - \left[H_{2,1}^\intercal \gamma \right]_+ \right) 1_{\hat{T}(H_{2,1}) \leq \lambda_n} \end{aligned} \quad (7)$$

where λ_n is a tuning parameter that we discuss in detail below. A lower bound $\mathcal{L}(c)$ can be defined by replacing the supremum with an infimum. The intuition behind this upper bound is as follows. Notice that the second term, $c^\intercal \hat{\Sigma}_1^{-1} \mathbb{P}_n B_1 \mathbb{U}_n$, in (6) is equal to

$$\begin{aligned} c^\intercal \hat{\Sigma}_1^{-1} \mathbb{P}_n B_1 \mathbb{U}_n 1_{\hat{T}(H_{2,1}) > \lambda_n} \\ + c^\intercal \hat{\Sigma}_1^{-1} \mathbb{P}_n B_1 \left(\left[H_{2,1}^\intercal (\mathbb{V}_n + \sqrt{n}\beta^*_{2,1}) \right]_+ - \left[H_{2,1}^\intercal \sqrt{n} \beta^*_{2,1} \right]_+ \right) 1_{\hat{T}(H_{2,1}) \leq \lambda_n}. \end{aligned} \quad (8)$$

The second term in (8) is algebraically equivalent to $c^\intercal \hat{\Sigma}_1^{-1} \mathbb{P}_n B_1 \mathbb{U}_n 1_{\hat{T}(H_{2,1}) \leq \lambda_n}$. However, we have rexpressed $\sqrt{n} H_{2,1}^\intercal \hat{\beta}_{2,1}$ as the sum of $H_{2,1}^\intercal \mathbb{V}_n = H_{2,1}^\intercal \sqrt{n}(\hat{\beta}_{2,1} - \beta^*_{2,1})$ and $H_{2,1}^\intercal \beta^*_{2,1} \sqrt{n}$; the latter quantity characterizes the degree of nonregularity of $\sqrt{n}(\hat{\beta}_1 - \beta^*_1)$ (see Theorem 2.1 below). Replacing $\sqrt{n}\beta^*_{2,1}$ with γ and taking the supremum over all $\gamma \in \mathbb{R}^{\dim(\beta^*_{2,1})}$ is one way of making the second term in (8) insensitive to local perturbations of $\beta^*_{2,1}$. More precisely, this yields a regular upper bound on the last term in (8). Combining this result with (6) yields (7). Theorem 2.1 below provides the asymptotic distribution of (7).

Suppose we want to construct a $(1 - \alpha) \times 100\%$ confidence interval for $c^\intercal \beta^*_1$. By con-

struction of $\mathcal{U}(c)$ and $\mathcal{L}(c)$ it follows that

$$c^\intercal\hat{\beta}_1 - \mathcal{U}(c)/\sqrt{n} \leq c^\intercal\beta_1^* \leq c^\intercal\hat{\beta}_1 - \mathcal{L}(c)/\sqrt{n}.$$

We approximate the distribution of the $\mathcal{U}(c)$ and $\mathcal{L}(c)$ using the bootstrap. Let $\hat{u}$ denote the $(1-\alpha/2) \times 100$ percentile of the bootstrap distribution of $\mathcal{U}(c)$, and let $\hat{l}$ denote the $(\alpha/2)\times 100$ percentile of the bootstrap distribution of $\mathcal{L}(c)$. Then $[c^\intercal\hat{\beta}_1 - \hat{u}/\sqrt{n},\, c^\intercal\hat{\beta}_1 - \hat{l}/\sqrt{n}]$ is the ACI for $c^\intercal\beta_1^*$.

Next we show that the ACI is asymptotically valid. First define

1. $B_t \triangleq (H_{t,0}^\intercal, H_{t,1}^T 1_{A_t=1})^\intercal$;
2. $\Sigma_{t,\infty} \triangleq PB_tB_t^\intercal$ for $t = 1, 2$;
3. $g_2(B_2, Y_2; \beta_2^*) \triangleq B_2(Y_2 - B_2^\intercal\beta_2^*)$;
4. $g_1(B_1, Y_1, H_2; \beta_1^*, \beta_2^*) \triangleq B_1\left(Y_1 + H_{2,0}^\intercal\beta_{2,0}^* + \left[H_{2,1}^\intercal\beta_{2,1}^*\right]_+ - B_1^\intercal\beta_1^*\right)$;

We use the following assumptions.

(A1) The histories H_2, features B_1, and outcomes Y_t, satisfy the moment inequalities $P||H_2||^2\,||B_1||^2 < \infty$ and $PY_2^2||B_2||^2 < \infty$.

(A2) The matrices $\Sigma_{t,\infty}$ and $\mathrm{Cov}\,(g_1, g_2)$ are strictly positive definite.

(A3) The sequence λ_n tends to infinity and satisfies $\lambda_n = o(n)$.

(A4) For $\gamma^* \in \mathbb{R}^{\dim(\beta_{2,1}^*)}$, there exists P_n a sequence of local alternatives converging to P in the sense that:

$$\int\left[\sqrt{n}\left(dP_n^{1/2} - dP^{1/2}\right) - \frac{1}{2}gdP^{1/2}\right]^2 \to 0,$$

for some measurable function g for which

– if $\beta^*_{2,n} \triangleq \arg\min_\beta P_n(Y_2 - Q_2(H_2, A_2; \beta)^2$, then $\beta^*_{2,1,n} \triangleq \beta^*_{2,1} + \gamma^*/\sqrt{n} + o(1/\sqrt{n})$ and

– $P_n||H_2||^2\,||B_1||^2$, $P_n Y_2^2 ||B_2||^2$ are bounded sequences.

Assumptions (A1)-(A2) are quite mild, requiring only full rank design matrices and some moment conditions. Requirement (A3) constrains a user-chosen tuning parameter and thus is always satisfied by appropriate choice of λ_n. Local alternatives provide a medium through which a glimpse of small sample behavior can be obtained, while retaining the mathematical convenience of large samples. Assumption (A4) facilitates a discussion of local alternatives without attempting to make the weakest possible assumptions (see van der Vaart and Wellner 1996, see also the remarks at the end of this section).

The first result regards the population upper bound $\mathcal{U}(c)$. Define $\tilde{Y}^*_{1,n} = H_{2,0}^\intercal \beta^*_{2,0,n} + \left[H_{2,1}^\intercal \beta^*_{2,1,n}\right]_+$ and $\beta^*_{1,n} \triangleq \arg\min_\beta P_n(\tilde{Y}^*_{1,n} - Q_1(H_1, A_1; \beta))^2$.

Theorem 2.1 (Validity of population bounds)**.** *Assume (A1)-(A3) and fix* $c \in \mathbb{R}^{\dim(\beta_1^*)}$.

1. $c^\intercal \sqrt{n}(\hat{\beta}_1 - \beta_1^*) \rightsquigarrow c^\intercal \mathbb{W}_\infty + c^\intercal \Sigma_{1,\infty}^{-1} P B_1 H_{2,1}^\intercal \mathbb{V}_\infty 1_{H_{2,1}^\intercal \beta^*_{2,1} > 0} + c^\intercal \Sigma_{1,\infty}^{-1} P B_1 \left[H_{2,1}^\intercal \mathbb{V}_\infty\right]_+ 1_{H_{2,1}^\intercal \beta^*_{2,1} = 0}$.

2. *If for each* n, *the underlying generative distribution is* P_n, *which satisfies (A4), then the limiting distribution of* $c^\intercal \sqrt{n}(\hat{\beta}_1 - \beta^*_{1,n})$ *is given by the distribution of*

$$c^\intercal \mathbb{W}_\infty + c^\intercal \Sigma_{1,\infty}^{-1} P B_1 H_{2,1}^\intercal \mathbb{V}_\infty 1_{H_{2,1}^\intercal \beta^*_{2,1} > 0} \\ + c^\intercal \Sigma_{1,\infty}^{-1} P B_1 \left(\left[H_{2,1}^\intercal (\mathbb{V}_\infty + \gamma^*)\right]_+ - \left[H_{2,1}^\intercal \gamma^*\right]_+\right) 1_{H_{2,1}^\intercal \beta^*_{2,1} = 0}. \quad (9)$$

3. *The limiting distribution of* $\mathcal{U}(c)$ *under both* P *and under* P_n *is equal to the distribution*

of

$$c^{\intercal}\mathbb{W}_{\infty} + c^{\intercal}\Sigma_{1,\infty}^{-1}PB_1H_{2,1}^{\intercal}\mathbb{V}_{\infty}1_{H_{2,1}^{\intercal}\beta_{2,1}^{*}>0}$$
$$+ \sup_{\gamma\in\mathbb{R}^{\dim(\beta_{2,1}^{*})}} c^{\intercal}\Sigma_{1,\infty}^{-1}PB_1\left(\left[H_{2,1}^{\intercal}(\mathbb{V}_{\infty}+\gamma)\right]_{+} - \left[H_{2,1}^{\intercal}\gamma\right]_{+}\right)1_{H_{2,1}^{\intercal}\beta_{2,1}^{*}=0}, \quad (10)$$

where $(\mathbb{W}_{\infty}^{\intercal},\ \mathbb{V}_{\infty}^{\intercal})$ *is asymptotically multivariate normal with mean zero.*

See the supplementary material for the proof and the formula for the $\mathrm{Cov}(\mathbb{W}_{\infty}, \mathbb{V}_{\infty})$. Notice that limiting distributions of $c^{\intercal}\sqrt{n}(\hat{\beta}_1 - \beta_1^{*})$ and $\mathcal{U}(c)$ (or equivalently $\mathcal{L}(c)$) are equal in the case $H_{2,1}^{\intercal}\beta_{2,1}^{*} \neq 0$ with probability one. That is, when there is a large treatment effect for almost all patients then the upper (or lower) bound is tight. However, when there is a non-null subset of patients for which there is no treatment effect, then the limiting distribution of the upper bound is stochastically larger than the limiting distribution of $c^{\intercal}\sqrt{n}(\hat{\beta}_1 - \beta_1^{*})$. Thus, the ACI adapts to the setting in which all patients experience a treatment effect.

Because the distribution of (9) depends on the local alternative, γ^{*}, $\hat{\beta}_1$ is a nonregular estimator (van der Vaart and Wellner, 1996). One might hope to construct an estimator of the distribution of (9) and use this estimator to approximate the distribution of $c^{\intercal}\sqrt{n}(\hat{\beta}_1 - \beta_1^{*})$. However a consistent estimator of the distribution of (9) does not exist because P_n is contiguous with respect to P (by assumption A4). To see this, let $F_{\gamma^{*}}(u)$ be the distribution of (9) evaluated at a point, u. If a consistent estimator, say $\hat{F}_n(u)$, existed, that is $\hat{F}_n(u)$ converges in probability to $F_{\gamma^{*}}(u)$ under P_n then the contiguity implies that $\hat{F}_n(u)$ converges in probability to $F_{\gamma^{*}}(u)$ under P. This is a contradiction (at best $\hat{F}_n(u)$ converges in probability to $F_0(u)$ under P). Because we can not consistently estimate γ^{*} and we do not know the value of γ^{*}, the tightest estimable upper bound on (9) is given by (10). As we shall next see, we are able to consistently estimate the distribution of (10).

In order to form confidence sets, the bootstrap distributions of $\mathcal{U}(c)$ and $\mathcal{L}(c)$ are used.

The next result regards the consistency of these bootstrap distributions. Let $\hat{\mathbb{P}}_n^{(b)}$ denote the bootstrap empirical measure, that is, $\hat{\mathbb{P}}_n^{(b)} \triangleq n^{-1}\sum_{i=1}^n M_{n,i}\delta_{\mathcal{T}_i}$ for $M_{n,1}, M_{n,2}, \ldots, M_{n,n} \sim$ Multinomial$(n,\, (1/n, 1/n, \ldots, 1/n))$. We use the superscript (b) to denote that a functional has been replaced by its bootstrap analogue, so that if $\omega \triangleq f(\mathbb{P}_n)$ then $w^{(b)} \triangleq f(\hat{\mathbb{P}}_n^{(b)})$. Denote the space of bounded Lipschitz-1 functions on $\mathbb{R}^2$ by $BL_1(\mathbb{R}^2)$. Furthermore, let $\mathbb{E}_M$ and P_M denote the expectation and probability with respect to the bootstrap weights. The following results are proved in the supplemental material.

Theorem 2.2. *Assume (A1)-(A3) and fix* $c \in \mathbb{R}^{\dim(\beta_1^*)}$. *Then* $(\mathcal{U}(c), \mathcal{L}(c))$ *and* $(\mathcal{U}^{(b)}(c), \mathcal{L}^{(b)}(c))$ *converge to the same limiting distribution in probability. That is,*

$$\sup_{v \in BL_1(\mathbb{R}^2)} \left| \mathbb{E}v\left((\mathcal{U}(c), \mathcal{L}(c))\right) - \mathbb{E}_M v\left(\left(\mathcal{U}^{(b)}(c), \mathcal{L}^{(b)}(c)\right)\right) \right|$$

converges in probability to zero.

Corollary 2.3. *Assume (A1)-(A3) and fix* $c \in \mathbb{R}^{\dim(\beta_1^*)}$. *Let* $\hat{u}$ *denote the* $(1-\alpha/2)\times 100$ *percentile of* $\mathcal{U}^{(b)}(c)$ *and* $\hat{l}$ *denote the* $(\alpha/2)\times 100$ *percentile of* $\mathcal{L}^{(b)}(c)$. *Then*

$$P_M\left(c^\intercal\hat{\beta}_1 - \hat{u}/\sqrt{n} \le c^\intercal\beta_1^* \le c^\intercal\hat{\beta}_1 - \hat{l}/\sqrt{n}\right) \ge 1 - \alpha + o_P(1).$$

Furthermore, if $P(H_{2,1}^\intercal\beta_{2,1}^* = 0) = 0$, *then the above inequality can be strengthened to equality.*

The preceding results show that the ACI can be use to construct valid confidence intervals regardless of the underlying parameters or generative model. Moreover, in settings where there is a treatment effect for almost every patient, the ACI delivers asymptotically exact coverage. See Section 4 for discussion of the choice of the tuning parameter λ_n.

Remark 2.4. The restriction on $\beta_{2,1,n}^*$ given in assumption (A4) is superfluous and can be seen to follow as a consequence of the convergence in quadratic mean condition. This is proved in the supplementary material.

Remark 2.5. Assumption (A2) can be relaxed if one is willing to proceed using generalized inverses. Since we are in a low dimensional setting we do not pursue this approach further.

3 Extending the ACI to many treatments

The two stage binary treatment setting which was addressed in the previous section provides the tools necessary to analyze data from many SMART trials including the ADHD study. However, there are a number of multistage randomized trials in which more than two treatments are available at each stage (Rush et al. 2003; Lieberman et al. 2005). In this section, we extend the ACI procedure for use with two stage trials with an arbitrary number of treatments available at each stage. The organization of this section parallels that of the previous section, however, the material is presented in a somewhat abbreviated fashion since much of the intuition has already been provided in earlier sections. In order to develop the results in this section, we require additional notation. Again, we observe trajectories $(X_1, A_1, Y_1, X_2, A_2, Y_2)$ drawn *i.i.d.* from some fixed and unknown distribution P. The treatment actions A_t take values in the set $\{1, \ldots, K_t\}$ for some fixed number of treatments K_t. In typical studies, K_t is no greater than five. We assume that the treatment action A_t is randomized with probabilities possibly depending on patient history, H_t ($H_t = \{X_1, A_1, \ldots, X_t\}$). We use the following linear model for the Q-function at time t:

$$Q_t(h_t, a_t; \beta_t) \triangleq \sum_{i=1}^{K_t} \beta_{t,i}^{\intercal} h_{t,1} 1_{a_t=i} \tag{11}$$

where as before $h_{t,1}$ is a vector of patient features constructed from the patient history, h_t and $\beta_t \triangleq (\beta_{t,1}^{\intercal}, \beta_{t,1}^{\intercal}, \ldots, \beta_{t,K_t}^{\intercal})^{\intercal}$. In (11) we omitted the main effect term (the term involving patient features that do not interact with treatment). This constraint permits compact theoretical expressions, but is unnecessary for the theoretical results. See the simulation

study for the use of a contrast coding. Note that according to this working model, if $h_{t,1}^{\intercal}\beta_{t,i} - \max_{j\neq i} h_{t,1}^{\intercal}\beta_{t,j} \approx 0$ for some $1 \leq i \leq K_t$, then at least two treatments are approximately optimal for a patient with history $H_{t,1} = h_{t,1}$. That is, there is *not* a unique best treatment for a patient with history $H_{t,1} = h_{t,1}$. Conversely, if $\min_{1\leq i\leq K_t} \left|h_{t,1}^{\intercal}\beta_{t,i} - \max_{j\neq i} h_{t,1}^{\intercal}\beta_{t,j}\right| \gg 0$, then *exactly* one treatment yields the best expected outcome for a patient with history $H_{t,1} = h_{t,1}$. As before, estimation of the optimal DTR is done using the Q-learning algorithm. The Q-learning algorithm proceeds as follows:

1. Regress Y_2 on H_2 and A_2 using (11) to obtain:

$$\hat{\beta}_2 \triangleq \arg\min_{\beta_2} \mathbb{P}_n(Y_2 - Q_2(H_2, A_2; \beta_2))^2,$$

   and subsequently the approximation $Q_2(h_2, a_2; \hat{\beta}_2)$ to the conditional mean $Q_2(h_2, a_2)$.

2. (a) Define the predicted future reward following the optimal policy as:

$$\begin{aligned} \tilde{Y}_1 &\triangleq Y_1 + \max_{a_2\in\{1,2,\ldots,K_2\}} Q_2(H_2, a_2; \hat{\beta}_2) &(12)\\ &= Y_1 + \max_{1\leq i\leq K_2} H_{2,1}^{\intercal}\hat{\beta}_{2,i} &(13)\end{aligned}$$

   (b) Regress $\tilde{Y}_1$ on H_1 and A_1 using (11) to obtain $\hat{\beta}_1 \triangleq \arg\min_{\beta_1} \mathbb{P}_n(\tilde{Y}_1 - Q_1(H_1, A_1; \beta_1))^2$.

3. Define the estimated optimal DTR $\hat{\pi} = (\hat{\pi}_1, \hat{\pi}_2)$ so that

$$\hat{\pi}_t(h_t) \triangleq \arg\max_{a_t\in\{1,2,\ldots,K_t\}} Q_t(h_t, a_t; \hat{\beta}_t).$$

As before, examination of the normal equations used to construct $\hat{\beta}_1$ combined with the definition of $\tilde{Y}_1$ show that $c^{\intercal}\sqrt{n}(\hat{\beta}_1 - \beta_1^*)$ can be decomposed as $c^{\intercal}\mathbb{W}_n + c^{\intercal}\hat{\Sigma}_1^{-1}\mathbb{P}_n B_1 \mathbb{U}_n$, where

the definitions of $\mathbb{W}_n$ and $\mathbb{U}_n$ have been generalized to

$$\begin{aligned}\mathbb{W}_n &= \hat{\Sigma}_1^{-1}\sqrt{n}\mathbb{P}_n B_1\left(Y_1+\left[\max_{1\le i\le K_2} H_{2,1}^{\intercal}\beta_{2,i}^{*}\right]-B_1^{\intercal}\beta_1^{*}\right),\\ \mathbb{U}_n &= \sqrt{n}\left(\left[\max_{1\le i\le K_2} H_{2,1}^{\intercal}\hat{\beta}_{2,i}\right]-\left[\max_{1\le i\le K_2} H_{2,1}^{\intercal}\beta_{2,i}^{*}\right]\right).\end{aligned}$$

The nonregularity of the limiting distribution of $c^{\intercal}\sqrt{n}(\hat{\beta}_1-\beta_1^{*})$ is apparent by noting the non-differentiable max operator in the definition of $\mathbb{U}_n$. Define $\mathbb{V}_{n,i}\triangleq\sqrt{n}(\hat{\beta}_{2,i}-\beta_{2,i}^{*})$ for $i=1,2,\ldots,K_2$.

The upper bound $\mathcal{U}(c)$ used to construct the ACI is given by

$$\begin{aligned}&c^{\intercal}\mathbb{W}_n+c^{\intercal}\hat{\Sigma}_1^{-1}\mathbb{P}_n B_1\mathbb{U}_n 1_{\min_i\hat{T}_i(h_{2,1})>\lambda_n}\\ &\quad+\sup_{\gamma} c^{\intercal}\hat{\Sigma}_1^{-1}\mathbb{P}_n B_1\left(\max_{1\le i\le K_2} H_{2,1}^{\intercal}\left(\mathbb{V}_{n,i}+\gamma_i\right)-\max_{1\le i\le K_2} H_{2,1}^{\intercal}\gamma_i\right)1_{\min_i\hat{T}_i(h_{2,1})\le\lambda_n},\end{aligned}\tag{14}$$

where $\gamma=(\gamma_1^{\intercal},\gamma_2^{\intercal},\ldots,\gamma_{K_2}^{\intercal})^{\intercal}$. The lower bound is formed similarly but with an infimum instead of a supremum. The test statistic, $\hat{T}_i(h_{2,1})$, is taken from the "multiple comparisons with the best" literature (see Hsu 1996 and references therein). This statistic is given by

$$\hat{T}_i(h_{2,1})\triangleq\frac{n\left(h_{2,1}^{\intercal}\hat{\beta}_{2,i}-\max_{j\neq i}h_{2,1}^{\intercal}\hat{\beta}_{2,j}\right)^2}{h_{2,1}^{\intercal}\hat{\zeta}_i h_{2,1}},$$

where $\hat{\zeta}_i$ is the usual plug-in estimator of $n\,\mathrm{Cov}(\hat{\beta}_{2,i}-\hat{\beta}_{2,j})$ for $j=\arg\max_{j\neq i}h_{2,1}^{\intercal}\hat{\beta}_{2,j}$, assuming the index j to be fixed a priori. Notice that $\min_i\hat{T}_i(h_{2,1})$ should be large if there is a uniquely optimal treatment for a patient with history $H_{2,1}=h_{2,1}$. On the other hand, $\hat{T}_i(h_{2,1})$ should be small if treatment i is the optimal treatment for a patient with history $h_{2,1}$ *and* there is more than one best treatment.

The theoretical results presented for the binary treatment ACI, including those regarding the bootstrap of the upper and lower bounds, hold in the many treatment case as well. While

there is no qualitative change in the required assumptions, they must however be generalized to accommodate an arbitrary number of treatments. The generalized assumptions along with statements of the theorems in the many treatment case can be found in the supplementary material (Supplement Section 2).

4 Empirical Study

In this section we contrast different choices of the potentially important tuning parameter λ_n and we provide an empirical evaluation of the ACI. Nine generative models are used in these evaluations; each of these generative models has two stages of treatment and two treatments at each stage. Generically, each of the models can be described as follows:

- $X_i \in \{-1, 1\}$, $A_i \in \{-1, 1\}$ for $i \in \{1, 2\}$
- $P(A_1 = 1) = P(A_1 = -1) = 0.5$, $P(A_2 = 1) = P(A_2 = -1) = 0.5$
- $X_1 \sim \text{Bernoulli}(0.5)$, $X_2 | X_1, A_1 \sim \text{Bernoulli}(\text{expit}(\delta_1 X_1 + \delta_2 A_1))$
- $Y_1 \triangleq 0$,

  $Y_2 = \gamma_1 + \gamma_2 X_1 + \gamma_3 A_1 + \gamma_4 X_1 A_1 + \gamma_5 A_2 + \gamma_6 X_2 A_2 + \gamma_7 A_1 A_2 + \epsilon$, $\epsilon \sim N(0, 1)$

where $\text{expit}(x) = \mathrm{e}^x/(1 + \mathrm{e}^x)$. This class is parameterized by nine values $\gamma_1, \gamma_2, ..., \gamma_7, \delta_1, \delta_2$. The analysis model uses patient feature vectors defined by:

$$\begin{aligned}
H_{2,0} &= (1, X_1, A_1, X_1 A_1, X_2)^\intercal \\
H_{2,1} &= (1, X_2, A_1)^\intercal \\
H_{1,0} &= (1, X_1)^\intercal \\
H_{1,1} &= (1, X_1)^\intercal.
\end{aligned}$$

Our analysis models are given by $Q_2(H_2, A_2; \beta_2) \triangleq H_{2,0}^{\intercal}\beta_{2,0} + H_{2,1}^{\intercal}\beta_{2,1}A_2$ and $Q_1(H_1, A_1; \beta_1) \triangleq H_{1,0}^{\intercal}\beta_{1,0} + H_{1,1}^{\intercal}\beta_{1,1}A_1$. We use contrast encoding for A_1 and A_2 to allow for a comparison with Chakraborty et al. (2009).

The form of this class of generative models is useful as it allows us to influence the degree of nonregularity present in our example problems through the choice of the γ_i and δ_i, and in turn evaluate performance in these different scenarios. Recall that in Q-learning, nonregularity occurs when more than one stage-two treatment produces nearly the same optimal expected reward for a set of patient histories that occur with positive probability. In the model class above, this occurs if the model generates histories for which $\gamma_5 A_2 + \gamma_6 X_2 A_2 + \gamma_7 A_1 A_2 \approx 0$, i.e., if it generates histories for which Q_2 depends weakly or not at all on A_2. By manipulating the values of γ_i and δ_i, we can control i) the probability of generating a patient history such that $\gamma_5 A_2 + \gamma_6 X_2 A_2 + \gamma_7 A_1 A_2 = 0$, and ii) the standardized effect size $E[(\gamma_5 + \gamma_6 X_2 + \gamma_7 A_1)/\sqrt{\mathrm{Var}(\gamma_5 + \gamma_6 X_2 + \gamma_7 A_1)}]$. Each of these quantities, denoted by p and ϕ, respectively, can be thought of as measures of nonregularity.

Table 1 provides the parameter settings; the first six settings were considered by Chakraborty et al. (2009), and are described by them as "nonregular", "near-nonregular", and "regular". To these six, we have added three additional examples labeled A, B, and C. Example A is an example of a strongly regular setting. Example B is an example of a nonregular setting where the nonregularity is strongly dependent on the stage 1 treatment action. In example B, for histories with $A_1 = 1$, there is a moderate effect of A_2 at the second stage. However, for histories with $A_1 = -1$, there is no effect of A_2 at the second stage, i.e., both actions at the second stage are equally optimal. In example C, for histories with $A_1 = 1$, there is a moderate effect of A_2, and for histories with $A_1 = -1$, there is a small effect of A_2. Thus example C is a 'near-nonregular' setting that behaves similarly to example B.

Example	γ	δ	Type	Regularity Measures	
1	$(0,0,0,0,0,0,0)^\intercal$	$(0.5,0.5)^\intercal$	nonregular	$p=1$	$\phi=0/0$
2	$(0,0,0,0,0.01,0,0)^\intercal$	$(0.5,0.5)^\intercal$	near-nonregular	$p=0$	$\phi=\infty$
3	$(0,0,-0.5,0,0.5,0,0.5)^\intercal$	$(0.5,0.5)^\intercal$	nonregular	$p=1/2$	$\phi=1.0$
4	$(0,0,-0.5,0,0.5,0,0.49)^\intercal$	$(0.5,0.5)^\intercal$	near-nonregular	$p=0$	$\phi=1.02$
5	$(0,0,-0.5,0,1.0,0.5,0.5)^\intercal$	$(1.0,0.0)^\intercal$	nonregular	$p=1/4$	$\phi=1.41$
6	$(0,0,-0.5,0,0.25,0.5,0.5)^\intercal$	$(0.1,0.1)^\intercal$	regular	$p=0$	$\phi=0.35$
A	$(0,0,-0.25,0,0.75,0.5,0.5)^\intercal$	$(0.1,0.1)^\intercal$	regular	$p=0$	$\phi=1.035$
B	$(0,0,0,0,0.25,0,0.25)^\intercal$	$(0,0)^\intercal$	nonregular	$p=1/2$	$\phi=1.00$
C	$(0,0,0,0,0.25,0,0.24)^\intercal$	$(0,0)^\intercal$	near-nonregular	$p=0$	$\phi=1.03$

Table 1: Parameters indexing the example models.

4.1 The choice of λ_n

We measure and compare the performance of four choices of the tuning parameter λ_n in terms of estimated coverage and average interval diameter. The intervals are constructed for intercept and the coefficient of the treatment indicator in the first stage Q-function in the nine generative models. We use a training set size of $n = 150$ in order to mimic the sample size of the ADHD study ($n = 138$). The online supplement contains a number of additional examples and sample sizes all displaying similar trends as presented here (Supplement Part V).

For the sequence λ_n we consider the following settings: $\lambda_n = \sqrt{\log\log n}$, $\log\log n$, $\log n$, $\sqrt{n}$, n. The intuition behind these settings is as follows. The supremum (infimum) used in the ACI can be thought of controlling the influence of committing a Type II error in the test of $\mathcal{N}_0(h_{2,1}) : h_{2,1}^\intercal\beta_{2,1}^* = 0$. On the other hand, the Type I error is controlled by the choice of λ_n. Recall that we reject the hypothesis $\mathcal{N}_0(h_{2,1})$ when $\hat{T}(h_{2,1}) > \lambda_n$. Thus, it is of interest to examine the (uniform) behaviour of $\hat{T}(h_{2,1})/\lambda_n$ across the set of $h_{2,1}$ for which $\mathcal{N}_0(h_{2,1})$ is true. Since the test statistic $\hat{T}$ is scale invariant (e.g. for any $\alpha > 0$ we have $\hat{T}(\alpha h_{2,1}) = \hat{T}(h_{2,1})$) is suffices to restrict our attention to unit vectors $h_{2,1}$ satisfying $\mathcal{N}_0(h_{2,1})$. We let $\mathcal{W} \triangleq \left\{h_{2,1} \in \mathbb{R}^{\dim(\beta_{2,1}^*)} : h_{2,1}^\intercal\beta_{2,1}^* = 0,\ ||h_{2,1}|| = 1\right\}$ denote these vectors of interest. Provided that λ_n tends to ∞ it follows that $\sup_{h\in\mathcal{W}}\hat{T}(h)/\lambda_n \to 0$ in probability. Further-

more, if λ_n grows faster than $\log\log n$ then the above convergence can be strengthened from in probability to almost surely using the law of the iterated logarithm (see Csorgo and Rosalsky 2003). However, consistency of the ACI also requires that $\lambda_n = o(n)$. Thus, $\lambda_n = n$ represents a rate that is too fast for consistency to hold; $\lambda_n = \log n$ is fast enough for strong (almost sure) control of the Type I error; $\lambda_n = \log\ \log n$ represents a rate that is at the boundary between almost sure and in convergence in probability; $\lambda_n = \sqrt{\log\log\ n}$ represents a rate that only ensures convergence in probability; $\lambda_n = \sqrt{n}$ represents a non-logarithmic rate that meets the consistency condition.

Tables (2) and (3) show the estimated coverage and interval diameter of the ACI across the five parameter settings for the nine generative models. The results appear stable across choices of λ_n for which the ACI is consistent. However, the ACI becomes quite conservative when λ_n is allowed to grow faster than log log n. Both in the simulation studies below as well as in the data analysis, we use $\lambda_n = \log\ \log\ n$.

4.2 An Evaluation of the ACI

We compare the empirical performance of the ACI with the centered percentile bootstrap (CPB), the soft-thresholding (ST) method of Chakraborty et al. (2009), and the simple plug-in pretesting estimator (PPE). The hard-thresholding of Moodie and Richardson (2007) is similar in theory and performance to the soft-thresholding approach; furthermore in orthogonal settings the lasso type penalization of Song et al. (2010) is equivalent to soft-thresholding, and so, Chakraborty's method is used to represent these alternate approaches.

The performance of each method is measured in terms of estimated coverage and interval diameter. We shall see that the ACI is conservative when there is no stage 2 treatment effect for all feature patterns; this is not unexpected since the ACI is based on the use of the upper/lower bounds. Despite the use of the bounds, ACI routinely delivers close to the nominal coverage and possesses competitive diameters. Competing methods fail to attain

$\beta_{1,1,1}$ $\lambda_n =$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B R	Ex. C R
$\sqrt{\log\log n}$	0.989	0.987	0.967	0.969	0.954	0.952	0.950	0.962	0.962
$\log\log n$	0.992	0.992	0.968	0.972	0.957	0.955	0.950	0.964	0.965
$\log n$	0.993	0.994	0.975	0.976	0.962	0.966	0.959	0.969	0.972
$\sqrt{n}$	0.994	0.995	0.975	0.976	0.967	0.972	0.968	0.973	0.975
n	0.994	0.995	0.975	0.976	0.969	0.972	0.968	0.975	0.976

$\beta_{1,0,1}$ $\lambda_n =$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B R	Ex. C R
$\sqrt{\log\log n}$	0.952	0.962	0.952	0.954	0.950	0.953	0.947	0.952	0.954
$\log\log n$	0.956	0.964	0.954	0.955	0.950	0.957	0.948	0.956	0.957
$\log n$	0.970	0.974	0.961	0.964	0.950	0.966	0.959	0.965	0.968
$\sqrt{n}$	0.971	0.975	0.963	0.968	0.954	0.973	0.965	0.974	0.978
n	0.971	0.975	0.987	0.987	0.979	0.980	0.975	0.983	0.984

Table 2: Monte Carlo estimates of coverage probabilities for the ACI methods at the 95% nominal level. Here, $\beta_{1,1,1}$ denotes the main effect of treatment and $\beta_{1,0,1}$ denotes the intercept. Estimates are constructed using 1000 datasets of size 150 drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models have two treatments at each of two stages. Examples are designated NR = nonregular, NNR = near-nonregular, R = regular.

$\beta_{1,1,1}$ $\lambda_n =$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B R	Ex. C R
$\sqrt{\log\log n}$	0.490	0.490	0.481	0.481	0.483	0.471	0.474	0.484	0.484
$\log\log n$	0.502	0.502	0.488	0.488	0.487	0.475	0.477	0.491	0.491
$\log n$	0.557	0.557	0.518	0.518	0.503	0.495	0.492	0.523	0.523
$\sqrt{n}$	0.583	0.582	0.533	0.533	0.513	0.514	0.511	0.540	0.540
n	0.586	0.586	0.538	0.538	0.525	0.521	0.519	0.543	0.543

$\beta_{1,0,1}$ $\lambda_n =$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B R	Ex. C R
$\sqrt{\log\log n}$	0.506	0.506	0.481	0.481	0.483	0.490	0.474	0.490	0.490
$\log\log n$	0.518	0.518	0.487	0.487	0.486	0.494	0.476	0.497	0.498
$\log n$	0.574	0.574	0.517	0.517	0.502	0.517	0.493	0.540	0.541
$\sqrt{n}$	0.596	0.596	0.536	0.536	0.515	0.543	0.519	0.571	0.572
n	0.598	0.598	0.576	0.576	0.565	0.586	0.565	0.579	0.579

Table 3: Monte Carlo estimates of mean width of the ACI method at the 95% nominal level. Here, $\beta_{1,1,1}$ denotes the main effect of treatment and $\beta_{1,0,1}$ denotes the intercept. Estimates are constructed using 1000 datasets of size 150 drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates with corresponding coverage significantly below 0.95 at the 0.05 level are marked with $*$. Models have two treatments at each of two stages. Examples are designated NR = nonregular, NNR = near-nonregular, R = regular.

nominal coverage on many of the examples.

Two additional alternative confidence sets, both proposed in Robins (2004), are the projection confidence set and a second confidence set that is based on a preliminary confidence set as opposed to the pretest considered here. Both proposals take advantage of the fact that the construction of a locally efficient score test for any fixed value of the vector $(\beta_1^\intercal, \beta_{2,1}^\intercal)$ is possible. In the first method a joint confidence set for $(\beta_1^{*\intercal}, \beta_{2,1}^{*}{}^\intercal)$ is constructed by inverting the locally efficient score test. Then this joint confidence set is projected to form a confidence set for $c^\intercal \beta_1^*$. Projection confidence intervals are generally conservative (Scheffe 1959; Nickerson 1994); that is, these confidence sets possess greater than the desired confidence level even when the problem is regular. As a result Robins proposes a second method that utilizes a "preliminary" confidence set. This preliminary $1-\epsilon$ joint confidence set is for $(\beta_{2,1}^* C^\perp \beta_1^*)$ where the columns of $C^\perp$ are orthogonal to c and the matrix $[c, C^\perp]$ is of full rank with number of columns equal to the dimension of β_1^*. For example this preliminary confidence set might be a projection of the confidence set for $(\beta_1^{*\intercal}, \beta_{2,1}^{*}{}^\intercal)$. Next assuming that $(\beta_{2,1}^* C^\perp \beta_1^*)$ is known to be $(\beta_{2,1}, C^\perp \beta_1)$, a locally efficient score test can be constructed for any fixed value of $c^\intercal \beta_1$. For each value of $(\beta_{2,1}, C^\perp \beta_1)$ in the preliminary confidence set, the locally efficient score test (at level α) is inverted. That is, the $1-\alpha-\epsilon$ asymptotic level confidence set contains all values of $c^\intercal \beta_1$ for which the locally efficient score test would accept for at least one value of $(\beta_{2,1}, C^\perp \beta_1)$ in the preliminary confidence set. These two approaches take advantage of the fact that if $\beta_{2,1}^*$ were known then inference for $c^\intercal \beta_1^*$ would be regular (thus the existence of the locally efficient score test).

To our knowledge neither method has been implemented (either in simulation or with data). Both pose difficult computational challenges that must be addressed prior to implementation. Both are nonconvex optimization problems and the projection confidence set may be the union of disjoint sets. As a result these two proposals are not evaluated here (see, however, the discussion for further comments).

We now briefly describe the three methods that we compare with the ACI. A natural first method to try is the bootstrap; thus the centered percentile bootstrap serves as a useful baseline for comparison. As discussed, the ST method works by shrinking the fitted regression $\hat{\beta}_1$ in the hopes of mitigating bias induced by nonregularity. In particular, for the working models we consider in this section; the ST estimators are:

$$\hat{\beta}_1^{ST} \triangleq \arg\min_{\beta_1} \mathbb{P}_n(\tilde{Y}_1^{ST} - B_1^\intercal \beta_1)^2 \tag{15}$$

$$\tilde{Y}_1^{ST} \triangleq Y_1 + H_{2,0}^\intercal \hat{\beta}_{2,0} + |H_{2,1}^\intercal \hat{\beta}_{2,1}| \left(1 - 3\frac{H_{2,1}^\intercal \hat{\Sigma}_{21,21} H_{2,1}}{n|H_{2,1}^\intercal \hat{\beta}_{2,1}|}\right)_+ . \tag{16}$$

In the above display, $\hat{\beta}_2$ and $\hat{\Sigma}_{21,21}$ are as described in previous sections. The constant 3 appearing in the ST method is motivated by an empirical Bayes interpretation of the thresholding (see work by Chakraborty et al. (2009) for more details). The form of the ST method shows that the modified predicted future reward following the optimal policy is shrunk most heavily when $h_{2,1}^\intercal \hat{\beta}_{2,1}$ is small. Which is to say, shrinkage occurs when there is little evidence that one treatment differs significantly from another for a patient with history $H_{2,1} = h_{2,1}$. The ST method is only developed for binary treatment.

The PPE confidence interval, in the two-stage binary treatment case, is formed by bootstrapping

$$c^\intercal \mathbb{W}_n + c^\intercal \hat{\Sigma}_1^{-1} \mathbb{P}_n B_1 \mathbb{U}_n 1_{\hat{T}(H_{2,1}) > \lambda_n} + c^\intercal \hat{\Sigma}_1^{-1} \mathbb{P}_n B_1 \left[H_{2,1}^\intercal \mathbb{V}_n\right]_+ 1_{\hat{T}(H_{2,1}) \le \lambda_n}. \tag{17}$$

This approach is natural as it partitions the data using a pretest and then uses a different estimator on each partition. A similar idea was employed by Chatterjee and Lahiri (2009) in their treatment of the Lasso. However, this approach is consistent under fixed but not local alternatives (see the supplemental material, Remark 1.9 and surrounding discussion for additional details); see also Leeb and Potscher 2005). As we will see below, this leads to

rather poor small sample performance. The primary reason for including this method is to motivate the importance of local alternatives and the utility of the supremum (infimum) in the construction of the ACI.

We first provide confidence intervals for the coefficient of A_1 (the treatment variable), $\beta^*_{1,1,0}$ in settings in which there are two or three treatments at stage 2. Note that given the working models and generative models defined by the parameter settings in Table 1, we can determine the exact value of any parameter $c^\intercal \beta^*_1$ of interest. The supplementary material (Supplement Section 4) contains confidence intervals for the treatment effect when $X_1 = 1$ (e.g. $\beta^*_{1,1,0} + \beta^*_{1,1,1}$). In addition, it contains estimated coverage probabilities and interval diameters for a range of sample sizes and a number of additional generative models, including those with three stages of treatment (Supplement Section 4).

Table 4 shows the estimated coverage for the coefficient of A_1, $\beta^*_{1,1,1}$. This simulation uses a sample size of 150, a total of 1000 Monte Carlo replications and 1000 bootstrap samples. Target coverage is .95. The CPB and PPE methods fare the worst in terms of coverage, each falling significantly below nominal coverage on fourteen of the eighteen examples, respectively. The ST method fails to cover for examples A,B and C. The reason for this under performance is that the ST method tends to over-shrink when treatment effects are larger as is the case in all of these examples. Recall that the ST method has not been developed for the setting in which there are more than two treatments at the second stage. The ACI delivers nominal coverage on all of the eighteen examples. The ACI is conservative on examples one and two. The average interval diameters are shown in Table 5. The ACI is the most conservative as is to be expected given that it is based on upper and lower bounds. However, the width is non-trivial.

The coefficient of A_1 is perhaps most relevant from a clinical perspective. However, from a methodological point of view, other contrasts can be illuminating. Table 6 shows the estimated coverage for the intercept using the same generative models described in the

preceding paragraph. The coverage of competing methods is quite poor collectively attaining nominal coverage on two examples. Particularly disturbing is that the ST method falls more than 30% below nominal levels. In contrast, the ACI delivers nominal coverage on all but two examples. Table 7 shows the average interval widths; the ACI is the widest but again is non-trivial.

5 Analysis of the ADHD study

In this section we illustrate the use of the ACI on data from the Adaptive Interventions for Children with ADHD study (Nahum-Shani et al. 2010a). The ADHD data we use here consists of $n = 138$ trajectories. These $n = 138$ trajectories are a subset of the original $N = 155$ observations. This subset was formed by removing the $N - n = 17$ patients that were either never randomized to an initial (first stage) treatment (14 patients), or had massive item missingness (3 patients). A description of each of the variables is described in Table (8). Notice that the outcomes Y_1 and Y_2 satisfy $Y_1 + Y_2 \equiv R$, where R is the teacher reported TIRS5 score at the last week of the study (week 32).

The first step in using Q-learning is to estimate a regression model for the second stage; this analysis only uses data from patients that were re-randomized during the 32 week study. Of the $n = 138$ patients, 79 were re-randomized before the study conclusion. The feature vectors at the second stage are $H_{2,0} \triangleq (1, X_{1,1}, X_{1,2}, X_{2,2}, X_{1,3}, X_{2,1}, A_1)^\intercal$ and $H_{2,1} \triangleq (1, X_{2,1}, A_1)^\intercal$. Thus, the Q-function $Q_2(H_2, A_2; \beta_2) \triangleq H_{2,0}^\intercal \beta_{2,0} + H_{2,1}^\intercal \beta_{2,1} A_2$ contains an interaction term between the second stage action A_2 and a patient's initial treatment A_1, an interaction between A_2 and adherence to their initial medication $X_{2,1}$, a main effect for A_2, and main effects for all the other terms. Table (9) provides the second stage least squares coefficients along with interval estimates. Examination of the residuals (not shown here) showed no obvious signs of model misspecification. In short, the linear model described

Two txts at stage 2	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.934*	0.935*	0.930*	0.933*	0.938	0.928*	0.939	0.925*	0.928*
PPE	0.931*	0.940	0.938	0.940	0.946	0.912*	0.931*	0.904*	0.903*
ST	0.948	0.945	0.938	0.942	0.952	0.943	0.919*	0.759*	0.762*
ACI	0.992	0.992	0.968	0.972	0.957	0.955	0.950	0.964	0.965

Three txts at stage 2	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.933*	0.938	0.915*	0.921*	0.931*	0.907*	0.940	0.885*	0.895*
PPE	0.931*	0.932*	0.927*	0.919*	0.932*	0.883*	0.918*	0.858*	0.856*
ACI	0.999	0.999	0.968	0.970	0.964	0.972	0.964	0.970	0.971

Table 4: Monte Carlo estimates of coverage probabilities of confidence intervals for the main effect of action, $\beta^*_{1,1,1}$ at the 95% nominal level. Estimates are constructed using 1000 datasets of size 150 drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. There is no ST method when there are three treatments at Stage 2. Examples are designated NR = nonregular, NNR = near-nonregular, R = regular.

Two txts at stage 2	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.385*	0.385*	0.430*	0.430*	0.457	0.436*	0.451	0.428*	0.428*
PPE	0.365*	0.366	0.419	0.419	0.452	0.418*	0.452*	0.404*	0.403*
ST	0.339	0.339	0.426	0.427	0.469	0.436	0.480*	0.426*	0.424*
ACI	0.502	0.502	0.488	0.488	0.487	0.475	0.477	0.491	0.491

Three txts at stage 2	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.446*	0.446	0.518*	0.518*	0.567*	0.518*	0.557	0.508*	0.507*
PPE	0.415*	0.415*	0.500*	0.500*	0.557*	0.486*	0.548*	0.467*	0.465*
ACI	0.716	0.716	0.663	0.663	0.643	0.643	0.625	0.673	0.673

Table 5: Monte Carlo estimates of the mean width of confidence intervals for the main effect of action $\beta^*_{1,1,1}$ at the 95% nominal level. Estimates are constructed using 1000 datasets of size 150 drawn from each model, and 1000 bootstraps drawn from each dataset. Models have two treatments at each of two stages. Widths with corresponding coverage significantly below nominal are marked with $*$. There is no ST method when there are three treatments at Stage 2. Examples are designated NR = nonregular, NNR = near-nonregular, R = regular.

Two txts at stage 2	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.892*	0.908*	0.924*	0.925*	0.940	0.930*	0.936	0.925*	0.931*
PPE	0.926*	0.930*	0.933*	0.934*	0.934*	0.907*	0.928*	0.910*	0.909*
ST	0.935*	0.930*	0.889*	0.878*	0.891*	0.620*	0.687*	0.686*	0.663*
ACI	0.956	0.964	0.954	0.955	0.950	0.957	0.948	0.956	0.957

Table 6: Monte Carlo estimates of coverage probabilities of confidence intervals for the coefficient of the intercept, $\beta^*_{1,0,1}$ at the 95% nominal level. Estimates are constructed using 1000 datasets of size 150 drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Examples are designated NR = nonregular, NNR = near-nonregular, R = regular.

Two txts at stage 2	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.404*	0.404*	0.430*	0.429*	0.457	0.449*	0.450	0.428*	0.428*
PPE	0.376*	0.376*	0.418*	0.418*	0.451*	0.448*	0.453*	0.410*	0.410*
ST	0.344*	0.344*	0.427*	0.427*	0.466*	0.469*	0.474*	0.430*	0.428*
ACI	0.518	0.518	0.487	0.487	0.486	0.494	0.476	0.497	0.498

Table 7: Monte Carlo estimates of the mean width of confidence intervals for the coefficient of the intercept, $\beta^*_{1,0,1}$ at the 95% nominal level. Estimates are constructed using 1000 datasets of size 150 drawn from each model, and 1000 bootstraps drawn from each dataset. Models have two treatments at each of two stages. Widths with corresponding coverage significantly below nominal are marked with $*$. Examples are designated NR = nonregular, NNR = near-nonregular, R = regular.

$X_{1,1} \in [0,3]$	:	baseline teacher reported mean ADHD symptom score. Measured at the end of the school year preceding the study.
$X_{1,2} \in \{0,1\}$	:	indicator of a diagnosis of ODD (oppositional defiant disorder) at baseline, coded so that $X_{1,3} = 0$ corresponds to no such diagnosis.
$X_{1,3} \in \{0,1\}$	:	indicator of a patient's prior exposure to ADHD medication, coded so that $X_{1,2} = 0$ corresponds to no prior exposure.
$A_1 \in \{-1,1\}$	:	initial treatment, coded so that $A_1 = -1$ corresponds to medication while $A_1 = 1$ corresponds to behavioral modification therapy.
$T \in \{6,7,\ldots 32\}$	:	right censored time in weeks until patient is re-randomized.
$Y_1 \triangleq R1_{T\geq 32}$	:	first stage response (see definition of R below).
$X_{2,1} \in \{0,1\}$	:	indicator of patient's adherence to their initial treatment. Adherence is coded so that a value of $X_{2,1} = 0$ corresponds to low adherence (taking less than 100% of prescribed medication or attending less than 75% of therapy sessions) while a value of $X_{2,1} = 1$ corresponds to high adherence.
$X_{2,2} \in \{1,8\}$	:	month of non-response.
$A_2 \in \{-1,1\}$	:	second stage treatment, coded so that $A_2 = -1$ corresponds to augmenting the initial treatment with the treatment *not* received initially, and $A_2 = 1$ corresponds to enhancing (increasing the dosage of) the initial treatment.
$R \in \{1,2,\ldots,5\}$	:	teacher reported Teacher Impairment Rating Scale (TIRS5) item score 32 weeks after initial randomization to treatment (Fabiano et al. 2006). The TIRS5 is coded so that higher values correspond to better clinical outcomes.
$Y_2 \triangleq R1_{T<32}$	:	second stage outcome.

Table 8: Components of a single trajectory in the ADHD study.

above seems to fit the data reasonably well.

Recall that the dependent variable in the first stage regression model is the predicted future outcome $\tilde{Y}_1 \triangleq Y_1 + \max_{a_2\in\{-1,1\}} Q_2(H_2, a_2; \hat{\beta}_2)$. Since the predictors used in the first stage must predate the assignment of first treatment, the available predictors in Table (8) are baseline ADHD symptoms $X_{1,1}$, diagnosis of ODD at baseline $X_{1,2}$, indicator of a patient's prior exposure to ADHD medication $X_{1,3}$, and first stage treatment A_1. The feature vectors for the second stage are $H_{1,0} \triangleq (1, X_{1,1}, X_{1,2}, X_{1,3})$ and $H_{1,1} \triangleq (1, X_{1,3})$, so that the first stage Q-function $Q_1(H_1, A_1; \beta_1) \triangleq H_{1,0}^{\intercal}\beta_{1,0} + H_{1,1}^{\intercal}\beta_{1,1}A_1$ contains an interaction term between the

Term	Coeff.	Estimate	Lower (5%)	Upper (95%)
Intercept	$\beta_{2,0,1}$	1.36	0.48	2.26
Baseline symptoms	$\beta_{2,0,2}$	0.94	0.48	1.39
ODD diagnosis	$\beta_{2,0,3}$	0.92	0.46	1.41
Month of non-response	$\beta_{2,0,4}$	0.02	-0.20	0.20
Prior Medication	$\beta_{2,0,5}$	-0.27	-0.77	0.21
Adherence	$\beta_{2,0,6}$	0.17	-0.28	0.66
First stage txt	$\beta_{2,0,7}$	0.03	-0.18	0.23
Second stage txt	$\beta_{2,1,1}$	-0.72	-1.13	-0.35
Second stage txt : Adherence	$\beta_{2,1,2}$	0.97	0.48	1.52
Second stage txt : First stage txt	$\beta_{2,1,3}$	0.05	-0.17	0.27

Table 9: Least squares coefficients and 90% interval estimates for second stage regression.

first stage action A_1 and a patient's prior exposure to ADHD medication $X_{1,3}$, a main effect for A_1, and main effects for all other covariates. The first stage regression coefficients are estimated using least squares $\hat{\beta}_1 \triangleq \arg\min_{\beta_1} \mathbb{P}_n(\tilde{Y}_1 - Q_1(H_1, A_1; \beta_1))^2$. Table (10) provides the least squares coefficients along with interval estimates formed using the ACI. Plots of the residuals for this model (not shown here) show no obvious signs of model misspecification. Again a linear model seems to provide a reasonable approximation to the Q-function in the first stage.

Term	Coeff.	Estimate	Lower (5%)	Upper (95%)
Intercept	$\beta_{1,0,1}$	2.61	2.09	3.08
Baseline symptoms	$\beta_{1,0,2}$	0.72	0.46	1.01
ODD diagnosis	$\beta_{1,0,3}$	0.75	0.38	1.07
Prior med. exposure	$\beta_{1,0,4}$	-0.37	-0.82	0.01
Initial txt	$\beta_{1,1,1}$	0.17	-0.05	0.39
Initial txt : Prior med. exposure	$\beta_{1,1,2}$	-0.32	-0.60	-0.05

Table 10: Least squares coefficients and 90% ACI interval estimates for first stage regression.

To construct an estimate of the optimal DTR, recall that for any $H_t = h_t,\ t = 1, 2$ the estimated optimal DTR $\hat{\pi} = (\hat{\pi}_1, \hat{\pi}_2)$ satisfies $\hat{\pi}_t(h_t) \in \arg\max_{a_t} Q(h_t, a_t; \hat{\beta}_t)$. The coefficients in Table (9) and the form of the second stage Q-function reveal that the second stage decision

rule $\hat{\pi}_2$ is quite simple. In particular, $\hat{\pi}_2$ prescribes treatment enhancement to patients with high adherence to their initial medication and it prescribes treatment augmentation to patients with low adherence to their initial medication. The first stage decision rule $\hat{\pi}_1$ is equally simplistic. The coefficients in Table (10) show that the first stage decision rule, $\hat{\pi}_1$ prescribes medication to patients who have had prior exposure to medication, and behavioral modification to patients who have not had any such prior exposure.

The prescriptions given by the estimated optimal DTR $\hat{\pi}$ are excessively decisive. That is, they recommend one and only one treatment regardless of the amount of evidence in the data to support that the recommended treatment is in fact optimal. When there is insufficient evidence to recommend a single treatment as best for a given patient history, it is preferred to leave the choice of treatment to the clinician. This allows the clinician to recommend treatment based on cost, local availability, patient individual preference, and clinical experience. One way to assess if there is sufficient evidence to recommend a unique optimal treatment for a patient is to construct a confidence interval for the predicted difference in mean response across treatments. In the case of binary treatments, for a fixed patient history $H_t = h_t$, one would construct a confidence interval for the difference $Q_t(h_t, 1; \beta_t^*) - Q_t(h_1, -1; \beta_t^*) = c^\intercal \beta_t^*$ where $c = (\mathbf{0}^\intercal, h_{t,1}^\intercal)^\intercal$. If this confidence interval contains zero then one would conclude that there is insufficient evidence at the nominal level for a unique best treatment.

In this example, the patient features that interact with treatment are categorical. Consequently, we can construct confidence intervals for the predicted difference in mean response across treatments for every possible patient history. These confidence intervals are given in table (11). The 90% confidence intervals suggest that there is insufficient evidence at the first stage to recommend a unique best treatment for each patient history. Rather, we would prefer not to make a strong recommendation at stage one, and leave treatment choice solely at the discretion of the clinician. Conversely, in the second stage, the 90% confidence intervals suggest that there is evidence to recommend a unique best treatment when a patient had

low adherence—knowledge that is important for evidence-based clinical decision making.

Stage	History	Contrast for $\beta_{t,1}$	Lower (5%)	Upper (95%)	Conclusion
1	Had prior med.	(1 1)	-0.48	0.16	Insufficient evidence
1	No prior med.	(1 0)	-0.05	0.39	Insufficient evidence
2	High adherence and BMOD	(1 1 1)	-0.08	0.69	Insufficient evidence
2	Low adherence and BMOD	(1 0 1)	-1.10	-0.28	Sufficient evidence
2	High adherence and MEDS	(1 1 -1)	-0.18	0.62	Insufficient evidence
2	Low adherence and MEDS	(1 0 -1)	-1.25	-0.29	Sufficient evidence

Table 11: Confidence intervals for the predicted difference in mean response across treatments across different patient histories at the 90% level. Confidence intervals that contain zero indicate insufficient evidence for recommending a unique best treatment.

6 Discussion

The task of constructing valid confidence intervals for the parameters in the Q-function is both scientifically important and statistically challenging. In this paper we offer a first step toward conducting inference in DTRs that is valid under local alternatives, computationally efficient, and easy to apply. The method presented here provides asymptotically valid intervals regardless of the true configuration of underlying parameters β_t^* or the joint distribution of patient histories H_t for $t = 1, 2, \ldots, T$. Theoretical guarantees were supported by a suite of test examples in which the ACI performed favorably to competitors. The ACI is conservative when all of the coefficients of terms involving the second stage treatment are zero. It is our experience that efforts to reduce this conservatism negatively impacts the performance of the resulting confidence interval for other generative models; we conjecture that this conservatism can not be ameliorated without negatively impacting the overall performance of

the confidence interval.

Robins (2004) second proposal for a confidence set (see Section 4.2 above for a description relevant to this setting), also adapts to regularity as is the case with the ACI. A critical difference between the ACI and Robins' second proposal is the way in which the pretest/preliminary confidence set is employed. Intuitively both are used to restrict the projection to a smaller set. However the ACI uses the pretest to bound only the nonsmooth parts of the statistic (the estimator of the stage 2 parameters are plugged-in the smooth parts of the statistic), whereas Robins' second proposal uses the preliminary confidence set to bound both the smooth and nonsmooth parts of the score statistic. This argues in favor of the ACI. On the other hand, in Robins' second proposal the confidence set is a union of acceptance regions and the preliminary confidence set is used to restrict this union whereas the ACI can be viewed as the acceptance region for a supremum statistic. This argues in favor of Robins' proposal. It would be most interesting to develop approximate algorithms that facilitate the use of Robins' second confidence set and then to compare the resulting approximate confidence set with the ACI.

There are a number of further avenues for work on this problem; we conclude by identifying three of the most interesting. The first is extending the ACI to problems where parameters are shared across stages. This setting occurs when a patient's status is modeled as series of renewals (as is often assumed in settings with a very large number of stages) or when smoothness across stages is assumed. The second area of interest is the so-called "large p small n" paradigm where the number of predictors in the Q-function far exceeds the number of observations. This setting arises, for example, when a patient's genetic information might be used to tailor treatment. A complication to the question of inference in this setting is that it is preceded by the more fundamental question of how one should even build Q-functions when $p \gg n$, but variable selection methods used in one-stage regression are likely to find use in the multi-stage case as well. Penalized estimation and Q-learning

in one-stage decision problems are discussed in (Qian and Murphy 2009) and in multi-stage problems in (Song et al., 2010). The third area of interest concerns reducing the bias in the estimation of the stage 1 treatment effect (recall that if the stage 2 effect is zero for a some patient features then the bias is of order $1/\sqrt{n}$). The most promising work in this area seems to be that of Song et al., (2010). This work produces nonregular estimators; it would be most interesting to develop confidence intervals that are valid under local alternatives even though the estimators are nonregular.

References

Donald W. Andrews and Gustavo Soares. Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection. *SSRN eLibrary*, 2007.

D.W.K. Andrews. Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica*, 69(3):683–734, 2001.

D.W.K. Andrews. Generalized method of moments estimation when a parameter is on a boundary. *Journal of Business and Economic Statistics*, 20(4):530–544, 2002.

D.W.K. Andrews and P. Guggenberger. Incorrect asymptotic size of subsampling procedures based on post-consistent model selection estimators. *Journal of Econometrics*, 152(1):19–27, 2009.

D.W.K. Andrews and W. Ploberger. Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica: Journal of the Econometric Society*, pages 1383–1414, 1994.

R. Beran. Diagnosing bootstrap success. *Annals of the Institute of Statistical Mathematics*, 49(1):1–24, 1997.

PJ Bickel, AJ Klaassen, Y. Ritov, and JA Wellner. *Efficient and adaptive inference in semi-parametric models*. Johns Hopkins University Press, Baltimore, 1993.

D. Blatt, Susan A. Murphy, and J. Zhu. A-learning for approximate planning. Technical Report 04-63, The Methodology Center, Pennsylvania State University, 2004.

B. Chakraborty, S. Murphy, and V. Strecher. Inference for non-regular parameters in optimal dynamic treatment regimes. *Statistical Methods in Medical Research*, 19(3), 2009.

A Chatterjee and S.N. Lahiri. Bootstrapping Lasso estimators. *Journal of the American Statistical Association*, 106(494):608–625, June 2011.

Jeesen Chen. Notes on the bias-variance trade-off phenomenon. *A Festschrift for Herman Rubin: Institute of Mathematical Statistics*, 45:207–217, 2004.

Xu Cheng. Robust confidence intervals in nonlinear regression under weak identification. *Job Market Paper*, 2008.

Sándor Csörgő and Andrew Rosalsky. A survey of limit laws for bootstrapped sums. *International Journal of Mathematics and Mathematical Statistics*, 45:2835–2861, 2003.

Donald W.K. Donald. Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica*, 69:683–734, 2001.

Hani Doss and Jayaram Sethuraman. The price of bias reduction when there is no unbiased estimate. *Annals of Statistics*, 17(1):440–442, 1989.

RB DAgostino. Departures from normality, tests for. *Encyclopedia of Statistical Sciences*, 2:315–324, 1982.

G.A. Fabiano, W.E. Pelham, D.A. Waschbusch, E.M. Gnagy, B.B. Lahey, A.M. Chronis, and L. Burrows-MacLean. A practical measure of impairment: Psychometric properties of the

impairment rating scale in samples of children with attention deficit hyperactivity disorder and two school-based samples. *Journal of Clinical Child and Adolescent Psychology*, 35: 369–385, 2006.

R. Henderson, P. Ansell, and D. Alshibani. Regret-Regression for Optimal Dynamic Treatment Regimes. *Biometrics*, 66(4), 2009.

Keisuke Hirano and Jack Porter. Impossibility results for nondifferentiable functionals. Mpra paper, University Library of Munich, Germany, 2009. URL `http://econpapers.repec.org/RePEc:pra:mprapa:15990`.

Michael R. Kosorok. *Introduction to empirical processes and semiparametric inference*. Springer, 2008.

P.W. Lavori and R. Dawson. A design for testing clinical strategies: biased adaptive within-subject randomization. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 163(1):29–38, 2000.

H. Leeb and B.M. Poetscher. The finite-sample distribution of post-model-selection estimators and uniform versus nonuniform approximations. *Econometric Theory*, 19(1):100–142, 2003.

H. Leeb and B.M. Pötscher. Model selection and inference: Facts and fiction. *Econometric Theory*, 21(1):21–59, 2005.

J.A. Lieberman, T.S. Stroup, J.P. McEvoy, M.S. Swartz, R.A. Rosenheck, D.O. Perkins, R.S.E. Keefe, S.M. Davis, C.E. Davis, B.D. Lebowitz, et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *The New England journal of medicine*, 353 (12):1209, 2005.

R.C. Liu and L.D. Brown. Nonexistence of informative unbiased estimators in singular problems. *The Annals of Statistics*, pages 1–13, 1993a.

Richard C. Liu and Lawrence D. Brown. Nonexistence of informative unbiased estimators in singular problems. *Annals of Statistics*, 21(1):1–13, 1993b.

E.E.M. Moodie, T.S. Richardson, and D.A. Stephens. Demystifying optimal dynamic treatment regimes. *Biometrics*, 63(2):447–455, 2007.

E.E.M. Moodie, T.S. Richardson, and D.A. Stephens. Estimating optimal dynamic regimes: Correcting bias under the null. *Biometrics*, 63(2):447–455, 2010.

S.A. Murphy and L.M. Collins. Customizing treatment to the patient: adaptive treatment strategies. *Drug and alcohol dependence*, 88(Suppl 2):S1, 2007.

Susan A. Murphy. Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society, Series B*, 65(2):331–366, 2003.

Susan A. Murphy. A generalization error for q-learning. *Journal of Machine Learning Research*, 6:1073–1097, Jul 2005.

I. Nahum-Shani, M. Qian, D. Almirall, W.E. Pelham, B. Gnagy, G. Fabiano, J. Waxmonsky, J. Yu, and S.A. Murphy. Experimental design and primary data analysis methods for comparing adaptive interventions. Technical Report 10-108, The Methodology Center, The Pennsylvania State University, 2010a.

I. Nahum-Shani, M. Qian, D. Almirall, W.E. Pelham, B. Gnagy, G. Fabiano, J. Waxmonsky, J. Yu, and S.A. Murphy. Q-learning: a data analysis method for constructing adaptive interventions. Technical Report 10-107, The Methodology Center, The Pennsylvania State University, 2010b.

David M. Nickerson. Construction of a conservative region from projections of an exact confidence region in multiple linear regression. *The American Statistician*, 48(2):120–124, 1994.

R.A. Olshen. The conditional level of the F-test. *Journal of the American Statistical Association*, 68(343):692–698, 1973.

W.E. Pelham Jr and G.A. Fabiano. Evidence-based psychosocial treatments for attention-deficit/hyperactivity disorder. *Journal of Clinical Child & Adolescent Psychology*, 37(1): 184–214, 2008.

S. Pliszka. AACAP Work Group on Quality Issues. Practice parameter for the assessment and treatment of children and adolescents with attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry*, 46(7):894–921, 2007.

B.M. Pötscher. Effects of model selection on inference. *Econometric Theory*, 7(2):163–185, 1991.

Min Qian and Susan A. Murphy. Performance Guarantees for Individualized Treatment Rules. Technical Report 498, Department of Statistics, University of Michigan, 2009.

J.M. Robins. Optimal structural nested models for optimal sequential decisions. In *Proceedings of the Second Seattle Symposium in Biostatistics: Analysis of Correlated Data*, 2004.

A.J. Rush, M. Trivedi, and M. Fava. Depression, IV: STAR* D treatment trial for depression. *American Journal of Psychiatry*, 160(2):237, 2003.

H. Scheffe. *The Analysis of Variance.* John Wiley, New York, 1959.

J. Shao. Bootstrap sample size in nonregular cases. *Proceedings of the American Mathematical Society*, 122(4):1251–1262, 1994.

R Song, W.. Wang, D. Zeng, and M. Kosorok. Penalized q-learning for dynamic treatment regimes. Technical Report arXiv:1108.5338v1, arxiv.org, 2011.

A.A. Tsiatis. *Semiparametric theory and missing data*. Springer Verlag, 2006.

C.J.C.H. Watkins and P. Dayan. Q-learning. *Machine learning*, 8(3):279–292, 1992.

Jr. William E. Pelham, Gregory Fabiano, James Waxmonsky, Andrew Greiner, Martin Hoffman, Susan Murphy, E. Michael Foster, Jihnhee Yu, Elizabeth Gnagy, Ira Bhatia, Jessica Verley, and Kate Tresco. Adaptive pharmacological and behavioral treatments for children with adhd: Sequencing, combining, and escalating doses. *Institute of Educational Sciences' Third Annual Research Conference, Washington, DC.*, 2008.

H. Working and H. Hotelling. Application of the theory of error to the interpretation of trends. *Journal of the American Statistical Association*, 25(165):73–85, 1929.

Y. Zhao, M.R. Kosorok, and D. Zeng. Reinforcement learning design for cancer clinical trials. *Statistics in Medicine*, 28:3294–3315, 2009.

Statistical Inference in Dynamic Treatment Regimes

Online Supplement

Eric B. Laber, Daniel J. Lizotte, Min Qian, and Susan A. Murphy

November 10, 2011

Contents

1 Proofs of results stated in the main body of the article

Throughout Sections 1 and 2, let K denote a sufficiently large positive constant that may vary from line to line. Let D_p denote the space of $p \times p$ symmetric positive definite matrices equipped with the spectral norm, and for any $k \in (0,1)$, let D_p^k denote the subset of D_p with members having eigenvalues in the range $[k, 1/k]$. For any class of real-valued functions $\mathcal{F}$, let $\rho_P(f) \triangleq (P(f - Pf)^2)^{1/2}$ denote the centered L_2-norm on $\mathcal{F}$, $l^\infty(\mathcal{F})$ denote the space of uniformly bounded real-valued functions on $\mathcal{F}$ equipped with the sup norm, and $C_b(\mathcal{F})$ denote the subspace of $l^\infty(\mathcal{F})$ of continuous and bounded functions from $\mathcal{F}$ into $\mathbb{R}$, respectively. Furthermore, let $\mathbb{G}_n \triangleq \sqrt{n}(\mathbb{P}_n - P)$, $\mathbb{G}_n^{(b)} \triangleq \sqrt{n}(\hat{\mathbb{P}}_n^{(b)} - \mathbb{P}_n)$, and P_M denote probability taken with respect to the bootstrap weights defining the bootstrap empirical measure, respectively.

1.1 Results for second stage parameters

In this section we will characterize the limiting distributions of the second stage parameters under fixed and local alternatives. We will also derive the limiting distribution of the bootstrap analog of the second stage parameters. For convenience, let $p_{t0} \triangleq \dim(\beta^*_{t,0})$, $p_{t1} \triangleq \dim(\beta^*_{t,1})$, and $p_t \triangleq \dim(\beta^*_t) = p_{t0} + p_{t1}$ for $t = 1, 2$.

Theorem 1.1. *Assume (A1) and (A2) and fix $a \in \mathbb{R}^{p_2}$, then*

1. $a^\intercal \sqrt{n}(\hat{\beta}_2 - \beta_2^*) \rightsquigarrow_P a^\intercal \Sigma_{2,\infty}^{-1} \mathbb{Z}_\infty$,
2. $a^\intercal \sqrt{n}(\hat{\beta}_2^{(b)} - \hat{\beta}_2) \rightsquigarrow_{P_M} a^\intercal \Sigma_{2,\infty}^{-1} \mathbb{Z}_\infty$ *in P-probability; and*
3. *if in addition (A4) holds,* $a^\intercal \sqrt{n}(\hat{\beta}_2 - \beta_{2,n}^*) \rightsquigarrow_{P_n} a^\intercal \Sigma_{2,\infty}^{-1} \mathbb{Z}_\infty$,

where $\mathbb{Z}_\infty$ is a mean zero normal random vector with covariance matrix $P[B_2 B_2^\intercal (Y_2 - B_2^\intercal \beta_2^*)^2]$.

Proof. Define the class of functions $\mathcal{F}_2$ as

$$\mathcal{F}_2 \triangleq \{f(b_2, y_2; a, \beta_2) \triangleq a^\intercal b_2(y_2 - b_2^\intercal \beta_2) \,:\, a, \beta_2 \in \mathbb{R}^{p_2},\, ||a|| \le K,\, ||\beta_2|| \le K\}, \tag{1}$$

and the function $w_2 : \mathcal{D}_{p_2} \times l^\infty(\mathcal{F}_2) \times \mathbb{R}^{p_2} \times \mathbb{R}^{p_2} \to \mathbb{R}$ as

$$w_2(\Sigma, \mu, \beta_2, a) \triangleq \mu\left(a^\intercal \Sigma^{-1} B_2(Y_2 - B_2^\intercal \beta_2)\right). \tag{2}$$

Since the estimated covariance matrices $\hat{\Sigma}_2 = \mathbb{P}_n B_2 B_2^\intercal$ and $\hat{\Sigma}_2^{(b)} = \hat{\mathbb{P}}_n^{(b)} B_2 B_2^\intercal$ are weakly consistent (by Lemma 1.3), we will avoid additional notation by assuming they are nonsingular for all n without loss of generality. Thus

$$\begin{aligned} &a^\intercal \sqrt{n}(\hat{\beta}_2 - \beta_2^*) = w_2(\hat{\Sigma}_2, \mathbb{G}_n, \beta_2^*, a), \quad a^\intercal \sqrt{n}(\hat{\beta}_2^{(b)} - \hat{\beta}_2) = w_2(\hat{\Sigma}_2^{(b)}, \mathbb{G}_n^{(b)}, \hat{\beta}_2, a), \\ &\text{and } a^\intercal \sqrt{n}(\hat{\beta}_2 - \beta_{2,n}^*) = w_2(\hat{\Sigma}_2, \sqrt{n}(\mathbb{P}_n - P_n), \beta_{2,n}^*, a). \end{aligned}$$

In addition, note that $a^\intercal \Sigma_2^{-1} \mathbb{Z}_\infty = w_2(\Sigma_2, \mathbb{G}_\infty, \beta_2^*, a)$, where $\mathbb{G}_\infty$ is a tight Gaussian process in $l^\infty(\mathcal{F}_2)$ with covariance function $Cov(\mathbb{G}_\infty f_1, \mathbb{G}_\infty f_2) = P(f_1 - Pf_1)(f_2 - Pf_2)$. Results 1 and 3 follow from Lemmas 1.2 - 1.5 and the continuous mapping theorem (Theorem 1.3.6 of van der Vaart and Wellner 1996). Result 2 follows from the bootstrap continuous mapping theorem (Theorem 10.8 of Kosorok 2008) together with Lemmas 1.2 - 1.6. □

Lemma 1.2. *Under (A1), the function w_2 defined in (2) is continuous at points in $D_{p_2} \times C_b(\mathcal{F}) \times \mathbb{R}^{p_2} \times \mathbb{R}^{p_2}$.*

Proof. Let $\epsilon > 0$ be arbitrary and let $(\Sigma, \mu, \beta_2, a)$ be an element of $D_{p_2} \times C_b(\mathcal{F}) \times \mathbb{R}^{p_2} \times \mathbb{R}^{p_2}$. In addition, let $(\Sigma', \mu', \beta_2', a')$ be an element of $D_{p_2} \times l^\infty(\mathcal{F}) \times \mathbb{R}^{p_2} \times \mathbb{R}^{p_2}$. From the form of $\mathcal{F}$ and the moment assumptions in (A1) we see that if $\Sigma - \Sigma'$, $a - a'$, and $\beta_2 - \beta_2'$ are small

then so must $\rho_P(f - f')$ be small, where

$$\begin{aligned} f(B_2, Y_2) &= a^\intercal \Sigma^{-1} B_2 (Y_2 - B_2^\intercal \beta_2), \\ f'(B_2, Y_2) &= a'^\intercal \Sigma'^{-1} B_2 (Y_2 - B_2^\intercal \beta_2'). \end{aligned}$$

In particular, we can choose $\delta > 0$ sufficiently small so that $||\Sigma - \Sigma'|| + ||a - a'|| + ||\beta_2 - \beta_2'|| < \delta$ implies that $\rho_P(f - f')$ is small enough to guarantee, by appeal to the continuity of μ, that $|\mu(f) - \mu(f')| \le \epsilon/2$. Finally, note that

$$\left| w_2(\Sigma, \mu, \beta_2, a) - w_2(\Sigma', \mu', \beta_2', a') \right| \le |\mu(f) - \mu(f')| + ||\mu - \mu'||_{\mathcal{F}_2}.$$

Let $\delta' = \min(\delta, \epsilon/2)$, then $||\Sigma - \Sigma'|| + ||\mu - \mu'||_{\mathcal{F}_2} + ||\beta_2 - \beta_2'|| + ||a - a'|| < \delta'$ implies that $|w_2(\Sigma, \mu, \beta_2, a) - w_2(\Sigma', \mu', \beta_2', a')| \le \epsilon$. Thus, the desired result is proved. $\square$

Having established the continuity of w_2 the next step will be to characterize the limiting behavior of $\beta_{2,n}^*$, $\hat{\beta}_2$, $\hat{\Sigma}_2$, $\hat{\Sigma}_2^{(b)}$, and the limiting distributions of $\mathbb{G}_n$, $\sqrt{n}(\mathbb{P}_n - P_n)$, and $\sqrt{n}(\hat{\mathbb{P}}_n^{(b)} - \mathbb{P}_n)$. These limits are established in a series of lemmas. Once this has been accomplished we will be able to apply the continuous mapping theorem to obtain the limiting distributions of $\sqrt{n}(\hat{\beta}_2 - \beta_2^*)$, $\sqrt{n}(\hat{\beta} - \beta_{2,n}^*)$, and $\sqrt{n}(\hat{\beta}_2^{(b)} - \hat{\beta}_2)$. Define $\Sigma_{2,n} \triangleq P_n B_2 B_2^\intercal$.

Lemma 1.3. *Assume (A1)-(A2), then $\hat{\Sigma}_2 \to_P \Sigma_2$ and $\hat{\Sigma}_2^{(b)} \to_{P_M} \Sigma_2$ in P-probability as $n \to \infty$. Furthermore, if (A4) holds, then $\hat{\Sigma}_2 \to_{P_n} \Sigma_2$ as $n \to \infty$.*

Proof. The first two claims follow from weak law of large numbers (Bickel and Freedman 1981; Csorgo and Rosalsky 2003). For the third claim, note that $\hat{\Sigma}_2 - \Sigma_2 = (\hat{\Sigma}_2 - \Sigma_{2,n}) + (\Sigma_{2,n} - \Sigma_2)$ and $\hat{\Sigma}_2 - \Sigma_{2,n} \to_{P_n} 0$ by law of large numbers. Below we show that $\Sigma_{2,n} \to \Sigma_2$. This will complete the proof.

let $c \in \mathbb{R}^{p_2}$ be arbitrary and define $\nu \triangleq c^\intercal B_2 B_2^\intercal c$. We will show that $\int \nu (dP_n - dP) = o(1)$.

First, note that

$$\int \nu(dP_n - dP) = \int \nu(dP_n^{1/2} + dP^{1/2})(dP_n^{1/2} - dP^{1/2}).$$

Furthermore, the absolute value of the foregoing expression is bounded above by

$$\int |\nu||(dP_n^{1/2} + dP^{1/2})|(dP_n^{1/2} - dP^{1/2}) \leq \sqrt{\int \nu^2(dP_n^{1/2} + dP^{1/2})^2}\sqrt{\int (dP_n^{1/2} - dP^{1/2})^2},$$

where the last inequality is simply Hölder's inequality. Next, note that owing to the inequality $(\sqrt{a} + \sqrt{b})^2 \leq 2a + 2b$ it follows that

$$\int \nu^2(dP_n^{1/2} + dP^{1/2})^2 \leq 2\int \nu^2 dP_n + 2\int \nu^2 dP = O(1),$$

by appeal to (A4). Now write

$$\int (dP_n^{1/2} - dP^{1/2})^2 = n^{-1}\left\{ \int \left(\sqrt{n}(dP_n^{1/2} - dP^{1/2}) - \frac{1}{2} g dP^{1/2}\right)^2 \right.$$
$$\left. - \frac{1}{4}\int g^2 dP + \sqrt{n}\int g dP^{1/2}(dP_n^{1/2} - dP^{1/2}) \right\}.$$

The right hand side of the preceding display is equal to

$$O(1/n) + n^{-1/2}\int g dP^{1/2}(dP_n^{1/2} - dP^{1/2}) \leq O(1/n) + n^{-1/2}\sqrt{\int g^2 dP}\sqrt{\int (dP_n^{1/2} - dP^{1/2})^2},$$

which is $o(1)$. Thus $\Sigma_{2,n} \to \Sigma_2$. □

Lemma 1.4. *Under (A1) and (A2), $\hat{\beta}_2 \to_P \beta_2^*$ as $n \to \infty$. If, in addition (A4) holds, then $\lim_{n\to\infty} \sqrt{n}(\beta_{2,n}^* - \beta_2^*) = \Sigma_{2,\infty}^{-1} PgB_2(Y_2 - B_2^\intercal \beta_2^*)$.*

Proof. $\hat{\beta}_2 \to_P \beta_2^*$ follows from weak law of large numbers and Slutsky's lemma.

Recall that $0 = P_n B_2(Y_2 - B_2^\intercal \beta_{2,n}^*)$ which we can write as

$$\sqrt{n}(P_n - P)B_2(Y_2 - B_2^\intercal \beta_2^*) - \Sigma_{2,n}\sqrt{n}(\beta_2^* - \beta_{2,n}^*),$$

so that for sufficiently large n it follows that $\sqrt{n}(\beta_{2,n}^* - \beta_2^*) = \Sigma_{2,n}^{-1}\sqrt{n}(P_n - P)B_2(Y_2 - B_2^\intercal \beta_2^*)$. By appeal to (A4) it follows that for any vector $a \in \mathbb{R}^{p_2}$ we have $\sup_n P_n(a^\intercal B_2(Y_2 - B_2^\intercal \beta_2^*))^2 < \infty$. Theorem 3.10.12 of van der Vaart and Wellner (1996) ensures that

$$\sqrt{n}(P_n - P)B_2(Y_2 - B_2^\intercal \beta_2^*) \rightarrow PgB_2(Y_2 - B_2^\intercal \beta_2^*)$$

as $n \rightarrow \infty$. This completes the proof. □

Lemma 1.5. *Assume (A1)-(A2), then*

1) $\mathbb{G}_n \rightsquigarrow_P \mathbb{G}_\infty$ *in* $l^\infty(\mathcal{F}_2)$*, where* $\mathcal{F}_2$ *is defined in (1), and* $\mathbb{G}_\infty$ *is a tight Gaussian process in* $l^\infty(\mathcal{F}_2)$ *with covariance function* $Cov(\mathbb{G}_\infty f_1, \mathbb{G}_\infty f_2) = P(f_1 - Pf_1)(f_2 - Pf_2)$*; and*

2) $\sup_{\omega \in BL_1} \left|\mathbb{E}_M \omega(\sqrt{n}(\hat{\mathbb{P}}_n^{(b)} - \mathbb{P}_n)) - \mathbb{E}\omega(\mathbb{G}_\infty)\right| \rightarrow_{P^*} 0$ *in* $l^\infty(\mathcal{F}_2)$.

If, in addition (A4) holds, then

3) $\sqrt{n}(\mathbb{P}_n - P_n) \rightsquigarrow_{P_n} \mathbb{G}_\infty$ *in* $l^\infty(\mathcal{F}_2)$.

Proof. First note that $\mathcal{F}_2$ is a subset of the pairwise product of the linear classes $\{a^\intercal b_2 : a \in \mathbb{R}^{p_2}\}$ and $\{y_2 - b_2^\intercal \beta_2 : \beta \in \mathbb{R}^{p_2}\}$ each of which is VC-subgraph of index no more than $p_2 + 1$ and P-measurable. Under (A1), the envelope of $\mathcal{F}_2$, $F_2(B_2, Y_2) = K||B_2||(|Y_2| + K||B_2||)$, is square integrable. This implies that $\mathcal{F}_2$ is P-Donsker, and 1) follows immediately. 2) follows from Theorem 3.6.1 of van der Vaart and Wellner (1996). For 3), note that from (A4) it follows that $\sup_f |P_n f|$ is a bounded sequence. The result follows from theorem 3.10.12 of van der Vaart and Wellner (1996). □

Lemma 1.6. *The space* $C_b(\mathcal{F}_2)$ *is a closed subset of* $l^\infty(\mathcal{F}_2)$ *and* $\mathrm{P}(\mathbb{G}_\infty \in C_b(\mathcal{F}_2)) = 1$.

Proof. Let $\{\mu_n\}_{n=1}^{\infty}$ be a convergent sequence of elements in $C_b(\mathcal{F}_2)$ and μ_0 the limiting element. Fir the first claim, we only need to show that $||\mu_0||_{\mathcal{F}_2} = \sup_{f\in\mathcal{F}_2}|\mu_0(f)|$ is bounded, and for any $f \in \mathcal{F}$ and $\epsilon > 0$, there exists some positive δ depending on f so that $|\mu_0(f') - \mu_0(f)| < \epsilon$ for all $f' \in \mathcal{F}_2$ and $\rho_P(f', f) < \delta$. The boundedness argument follows by noticing that $||\mu_0||_{\mathcal{F}_2} \leq ||\mu_n||_{\mathcal{F}_2} + ||\mu_n - \mu_0||_{\mathcal{F}_2}$ for any n; in particular, for some fixed large enough n, $||\mu_n||_{\mathcal{F}_2}$ is bounded by the fact $\mu_n \in C_b(\mathcal{F}_2)$, and $||\mu_n - \mu_0||_{\mathcal{F}_2}$ is bounded above by a constant due to the convergence of μ_n to μ_0. For continuity, note that since μ_n converges to μ_0, we can choose n^* so that $||\mu_n - \mu_0|| < \epsilon/4$ for all $n \geq n^*$. In addition, by the continuity of μ_{n^*}, there exists some $\delta > 0$ so that $|\mu_{n^*}(f') - \mu_{n^*}(f)| < \epsilon$ for all $\rho_P(f', f) < \delta$. Thus

$$\begin{aligned} |\mu_0(f') - \mu_0(f)| &\leq |\mu_0(f) - \mu_{n^*}(f)| + |\mu_{n^*}(f') - \mu_0(f')| + |\mu_{n^*}(f) - \mu_{n^*}(f')| \\ &\leq 2||\mu_0 - \mu_{n^*}||_{\mathcal{F}_2} + |\mu_{n^*}(f) - \mu_{n^*}(f')| \\ &\leq 3\epsilon/4. \end{aligned}$$

This implies that $C_b(\mathcal{F})$ is closed.

Next note that $\mathbb{G}_\infty$ is a tight Gaussian process in $l^\infty(\mathcal{F}_2)$. By the argument in section 1.5 of van de Vaart and Wellner (1996), almost all sample paths $f \to \mathbb{G}_\infty(f, \omega)$ are uniformly ρ_2-continuous, where $\rho_2(f_1, f_2) = [P(\mathbb{G}_\infty f_1 - \mathbb{G}_\infty f_2)^2]^{1/2}$ is a semimetric on $\mathcal{F}$. Since $\rho_2(f_1, f_2) = [Var(f_1 - f_2)]^{1/2} \leq \rho_P(f_1, f_2)$, the continuity of the sample paths of $\mathbb{G}_\infty$ follows immediately. $\square$

1.2 A characterization of the first stage coefficients and the upper bound $\mathcal{U}(c)$

In this section we present the proofs for Theorem 2.1 and 2.2. We first derive an expansion for the first stage coefficients and two useful expansions of the upper bound $\mathcal{U}(c)$. The terms

in the forementioned expansions will be treated individually in subsequent sections. We will make use of the following functions.

1. $w_{11} : D_{p_1} \times D_{p_1 \times p_{20}} \times l^\infty(\mathcal{F}_{11}) \times l^\infty(\mathcal{F}_{11}) \times \mathbb{R}^{p_2} \times \mathbb{R}^{p_1+p_2} \to \mathbb{R}$ is defined as

$$\begin{aligned} w_{11}(\Sigma_1, \Sigma_{12}, \mu, \omega, \nu, \beta) &\triangleq \mu\left[c^\intercal \Sigma_1^{-1} B_1 \big(Y_1 + H_{2,0}^\intercal \beta_{2,0} + \left[H_{2,1}^\intercal \beta_{2,1}\right]_+ - B_1^\intercal \beta_1\big)\right] \\ &\quad + c^\intercal \Sigma_1^{-1} \Sigma_{12} \nu_0 + \omega\left(c^\intercal \Sigma_1^{-1} B_1 H_{2,1}^\intercal \nu_1 1_{H_{2,1}^\intercal \beta_{2,1}^* > 0}\right), \end{aligned} \tag{3}$$

where $D_{p_1 \times p_{20}}$ is the space of $p_1 \times p_{20}$ matrices equipped with the spectral norm, and $\mathcal{F}_{11} = \Big\{ f(b_1, y_1, h_{2,0}, h_{2,1}) = a_1^\intercal b_1 \big(y_1 + h_{2,0}^\intercal \beta_{2,0} + \left[h_{2,1}^\intercal \beta_{2,1}\right]_+ - b_1^\intercal \beta_1\big) + a_2^\intercal b_1 (h_{2,1}^\intercal \nu_1) 1_{h_{2,1}^\intercal \beta_{2,1}^* > 0} :$ $\beta = (\beta_1^\intercal, \beta_{2,0}^\intercal, \beta_{2,1}^\intercal)^\intercal \in \mathbb{R}^{p_1+p_2}, \nu = (\nu_0^\intercal, \nu_1^\intercal)^\intercal \in \mathbb{R}^{p_2}, a_1, a_2 \in \mathbb{R}^{p_1}, \max\{||a_1||, ||a_2||, ||\beta||, ||\nu||\} \leq K \Big\}$.

2. $w_{12} : D_{p_1} \times l^\infty(\mathcal{F}_{12}) \times \mathbb{R}^{p_{21}} \times \mathbb{R}^{p_{21}} \to \mathbb{R}$ is defined as

$$w_{12}(\Sigma_1, \mu, \nu, \gamma) \triangleq \mu\left[c^\intercal \Sigma_1^{-1} B_1 \left(\left[H_{2,1}^\intercal \nu + H_{2,1}^\intercal \gamma\right]_+ - \left[H_{2,1}^\intercal \gamma\right]_+\right) 1_{H_{2,1}^\intercal \beta_{2,1}^* = 0}\right], \tag{4}$$

where $\mathcal{F}_{12} = \Big\{ f(b_1, h_{2,1}) = a^\intercal b_1 \left([h_{2,1}^T \nu + h_{2,1}^\intercal \gamma]_+ - [h_{2,1}^\intercal \gamma]_+\right) 1_{h_{2,1}^\intercal \beta_{2,1}^* = 0} : a \in \mathbb{R}^{p_1}, \gamma,\ \nu \in \mathbb{R}^{p_{21}}, \max\{||a||, ||\nu||\} \leq K \Big\}$.

3. $\rho_{11} : D_{p_1} \times D_{p_{21}}^k \times l^\infty(\widetilde{\mathcal{F}}_{11}) \times \mathbb{R}^{p_{21}} \times \mathbb{R}^{p_{21}} \times \mathbb{R}^{p_{21}} \times \mathbb{R} \to \mathbb{R}$, is defined as

$$\begin{aligned} \rho_{11}(\Sigma_1, \Sigma_{21,21}, \mu, \nu, \eta, \gamma, \lambda) \triangleq \mu\Bigg[& c^\intercal \Sigma_1^{-1} B_1 \left(\left[H_{2,1}^\intercal \nu + H_{2,1}^\intercal \gamma\right]_+ - [H_{2,1}^\intercal \gamma]_+\right) \\ & \times \left(1_{\frac{(H_{2,1}^\intercal \nu + H_{2,1}^\intercal \eta)^2}{H_{2,1}^\intercal \Sigma_{21,21} H_{2,1}} \leq \lambda} - 1_{H_{2,1}^\intercal \beta_{2,1}^* = 0}\right)\Bigg], \end{aligned} \tag{5}$$

where $\widetilde{\mathcal{F}}_{11} = \Big\{ f(b_1, h_{2,1}) = a^\intercal b_1 \left([h_{2,1}^\intercal \nu - h_{2,1}^\intercal \gamma]_+ - [h_{2,1}^\intercal \gamma]_+\right) (1_{\frac{(h_{2,1}^\intercal \nu + h_{2,1}^\intercal \eta)^2}{h_{2,1}^\intercal \Sigma_{21,21} h_{2,1}} \leq \lambda} - 1_{h_{2,1}^\intercal \beta_{2,1}^* = 0}), :$ $a \in \mathbb{R}^{p_1}, \nu, \eta, \gamma \in \mathbb{R}^{p_{21}}, \max\{||a||, ||\nu||\} \leq K, \lambda \in \mathbb{R}, \Sigma_{21,21} \in D_{p_{21}}^k \Big\}$.

4. $\rho_{12} : D_{p_1} \times l^{\infty}(\widetilde{\mathcal{F}}_{12}) \times \mathbb{R}^{p_{21}} \times \mathbb{R}^{p_{21}} \to \mathbb{R}$, defined as

$$\rho_{12}(\Sigma_1, \mu, \nu, \eta) \triangleq \mu\Big[c^{\intercal}\Sigma_1^{-1}B_1\left(\left[H_{2,1}^{\intercal}\nu + H_{2,1}^{\intercal}\eta\right]_+ - [H_{2,1}^{\intercal}\eta]_+ - H_{2,1}^{\intercal}\nu\right)1_{H_{2,1}^{\intercal}\beta_{2,1}^*>0}$$
$$+ c^{\intercal}\Sigma_1^{-1}B_1\left(\left[H_{2,1}^{\intercal}\nu + H_{2,1}^{\intercal}\eta\right]_+ - [H_{2,1}^{\intercal}\eta]_+\right)1_{H_{2,1}^{\intercal}\beta_{2,1}^*<0}\Big], \quad (6)$$

where $\widetilde{\mathcal{F}}_{12} = \Big\{a^{\intercal}b_1\big(\left[h_{2,1}^{\intercal}\nu + h_{2,1}^{\intercal}\eta\right]_+ - [h_{2,1}^{\intercal}\eta]_+ - h_{2,1}^{\intercal}\nu\big)1_{h_{2,1}^{\intercal}\beta_{2,1}^*>0} - a^{\intercal}b_1\big(\left[h_{2,1}^{\intercal}\nu + h_{2,1}^{\intercal}\eta\right]_+ - [h_{2,1}^{\intercal}\eta]_+\big)1_{h_{2,1}^{\intercal}\beta_{2,1}^*<0} : a \in \mathbb{R}^{p_1}, \nu \in \mathbb{R}^{p_{21}}, \max\{||a||, ||\nu||\} \leq K, \eta \in \mathbb{R}^{p_{21}}\Big\}$.

Using the foregoing functions, we have the following expressions for the first stage parameters:

$$\begin{aligned}
c^{\intercal}\sqrt{n}(\hat{\beta}_1 - \beta_1^*) &= w_{11}(\hat{\Sigma}_1, \hat{\Sigma}_{12}, \mathbb{G}_n, \mathbb{P}_n, \sqrt{n}(\hat{\beta}_2 - \beta_2^*), (\beta_1^{*\intercal}, \beta_2^{*\intercal})^{\intercal}) \\
&\quad + w_{12}(\hat{\Sigma}_1, \mathbb{P}_n, \sqrt{n}(\hat{\beta}_{2,1} - \beta_{2,1}^*), \sqrt{n}\beta_{2,1}^*) \\
&\quad + \rho_{12}(\hat{\Sigma}_1, \mathbb{P}_n, \sqrt{n}(\hat{\beta}_{2,1} - \beta_{2,1}^*), \sqrt{n}\beta_{2,1}^*); && (7) \\
\sqrt{n}(\hat{\beta}_1 - \beta_{1,n}^*) &= w_{11}(\hat{\Sigma}_1, \hat{\Sigma}_{12}, \sqrt{n}(\mathbb{P}_n - P_n), \mathbb{P}_n, \sqrt{n}(\hat{\beta}_2 - \beta_{2,n}^*), (\beta_{1,n}^{*\intercal}, \beta_{2,n}^{*\intercal})^{\intercal}) \\
&\quad + w_{12}(\hat{\Sigma}_1, \mathbb{P}_n, \sqrt{n}(\hat{\beta}_{2,1} - \beta_{2,1,n}^*), \sqrt{n}\beta_{2,1,n}^*) \\
&\quad + \rho_{12}(\hat{\Sigma}_1, \mathbb{P}_n, \sqrt{n}(\hat{\beta}_{2,1} - \beta_{2,1,n}^*), \sqrt{n}\beta_{2,1,n}^*), && (8)
\end{aligned}$$

where $\hat{\Sigma}_{12} = \mathbb{P}_n B_1 H_{2,0}^{\intercal}$. Similarly, we can express the upper bound $\mathcal{U}(c)$ in terms of the above functions:

$$\begin{aligned}
\mathcal{U}(c) &= w_{11}(\hat{\Sigma}_1, \hat{\Sigma}_{12}, \mathbb{G}_n, \mathbb{P}_n, \sqrt{n}(\hat{\beta}_2 - \beta_2^*), (\beta_1^{*\intercal}, \beta_2^{*\intercal})^{\intercal}) \\
&\quad + \rho_{12}(\hat{\Sigma}_1, \mathbb{P}_n, \sqrt{n}(\hat{\beta}_{2,1} - \beta_{2,1}^*), \sqrt{n}\beta_{2,1}^*) \\
&\quad - \rho_{11}(\hat{\Sigma}_1, \hat{\Sigma}_{21,21}, \mathbb{P}_n, \sqrt{n}(\hat{\beta}_{2,1} - \beta_{2,1}^*), \sqrt{n}\beta_{2,1}^*, \sqrt{n}\beta_{2,1}^*, \lambda_n) \\
&\quad + \sup_{\gamma \in \mathbb{R}^{p_{2,1}}} \Big\{ w_{12}(\hat{\Sigma}_1, \mathbb{P}_n, \sqrt{n}(\hat{\beta}_{2,1} - \beta_{2,1}^*), \gamma) \\
&\qquad + \rho_{11}(\hat{\Sigma}_1, \hat{\Sigma}_{21,21}, \mathbb{P}_n, \sqrt{n}(\hat{\beta}_{2,1} - \beta_{2,1}^*), \sqrt{n}\beta_{2,1}^*, \gamma, \lambda_n) \Big\}. && (9)
\end{aligned}$$

We will also make use of the following alternative expression for the upper bound $\mathcal{U}(c)$ under P_n:

$$\begin{aligned}
\mathcal{U}(c) \;=\;& w_{11}(\hat{\Sigma}_1, \hat{\Sigma}_{12}, \sqrt{n}(\mathbb{P}_n - P_n), \mathbb{P}_n, \sqrt{n}(\hat{\beta}_2 - \beta^*_{2,n}), (\beta^{*\intercal}_{1,n}, \beta^{*\intercal}_{2,n})^\intercal) \\
&+ \rho_{12}(\hat{\Sigma}_1, \mathbb{P}_n, \sqrt{n}(\hat{\beta}_{2,1} - \beta^*_{2,1,n}), \sqrt{n}\beta^*_{2,1,n}) \\
&- \rho_{11}(\hat{\Sigma}_1, \hat{\Sigma}_{21,21}, \mathbb{P}_n, \sqrt{n}(\hat{\beta}_{2,1} - \beta^*_{2,1,n}), \sqrt{n}\beta^*_{2,1,n}, \sqrt{n}\beta^*_{2,1,n}, \lambda_n) \\
&+ \sup_{\gamma \in \mathbb{R}^{p_{21}}} \Big\{ w_{12}(\hat{\Sigma}_1, \mathbb{P}_n, \sqrt{n}(\hat{\beta}_{2,1} - \beta^*_{2,1,n}), \gamma) \\
&\quad + \rho_{11}(\hat{\Sigma}_1, \hat{\Sigma}_{21,21}, \mathbb{P}_n, \sqrt{n}(\hat{\beta}_{2,1} - \beta^*_{2,1,n}), \sqrt{n}\beta^*_{2,1,n}, \gamma, \lambda_n) \Big\}. \qquad (10)
\end{aligned}$$

Similarly, we will make use of following expression for the bootstrap analog of the upper bound:

$$\begin{aligned}
\hat{\mathcal{U}}^{(b)}(c) \;=\;& w_{11}(\hat{\Sigma}^{(b)}_1, \hat{\Sigma}^{(b)}_{12}, \sqrt{n}(\mathbb{P}^{(b)}_n - \mathbb{P}_n), \mathbb{P}^{(b)}_n, \sqrt{n}(\hat{\beta}^{(b)}_2 - \hat{\beta}_2), (\hat{\beta}^\intercal_1, \hat{\beta}^\intercal_2)^\intercal) \\
&+ \rho_{12}(\hat{\Sigma}^{(b)}_1, \hat{\mathbb{P}}^{(b)}_n, \sqrt{n}(\hat{\beta}^{(b)}_{2,1} - \hat{\beta}_{2,1}), \sqrt{n}\hat{\beta}_{2,1}) \\
&- \rho_{11}(\hat{\Sigma}^{(b)}_1, \hat{\Sigma}^{(b)}_{21,21}, \hat{\mathbb{P}}^{(b)}_n, \sqrt{n}(\hat{\beta}^{(b)}_{2,1} - \hat{\beta}_{2,1}), \sqrt{n}\hat{\beta}_{2,1}, \sqrt{n}\hat{\beta}_{2,1}, \lambda_n) \\
&+ \sup_{\gamma \in \mathbb{R}^{p_{21}}} \Big\{ w_{12}(\hat{\Sigma}^{(b)}_1, \hat{\mathbb{P}}^{(b)}_n, \sqrt{n}(\hat{\beta}^{(b)}_{2,1} - \hat{\beta}_{2,1}), \gamma) \\
&\quad + \rho_{11}(\hat{\Sigma}^{(b)}_1, \hat{\Sigma}^{(b)}_{21,21}, \hat{\mathbb{P}}^{(b)}_n, \sqrt{n}(\hat{\beta}^{(b)}_{2,1} - \hat{\beta}_{2,1}), \sqrt{n}\hat{\beta}_{2,1}, \gamma, \lambda_n) \Big\}. \qquad (11)
\end{aligned}$$

Below we argue that ρ_{11} and ρ_{12} are negligible and w_{11} and w_{12} are continuous in such a fashion so as to facilitate the use of continuous mapping theorems as presented in the previous section.

Analysis of Error Terms ρ_{11} and ρ_{12}

In this section, we show that the functions ρ_{11} and ρ_{12} in the expressions for the first stage parameters and upper bounds are negligible. The function ρ_{11} is more difficult to handle so

we address it here and omit the proof involving ρ_{12} as it involves similar ideas.

Theorem 1.7. *Assume (A1)-(A3). Then*

1. $\sup_{\gamma\in\mathbb{R}^{p_{21}}} \left|\rho_{11}(\hat{\Sigma}_1, \hat{\Sigma}_{21,21}, \mathbb{P}_n, \sqrt{n}(\hat{\beta}_{2,1} - \beta^*_{2,1}), \sqrt{n}\beta^*_{2,1}, \gamma, \lambda_n)\right| \to_P 0$, *and*

2. $\sup_{\gamma\in\mathbb{R}^{p_{21}}} \left|\rho_{11}(\hat{\Sigma}^{(b)}_1, \hat{\Sigma}^{(b)}_{21,21}, \hat{\mathbb{P}}^{(b)}_n, \sqrt{n}(\hat{\beta}^{(b)}_{2,1} - \hat{\beta}_{2,1}), \sqrt{n}\hat{\beta}_{2,1}, \gamma, \lambda_n)\right| \to_{P_M} 0$ *almost surely P.*

If, in addition, we assume (A4), then

3. $\sup_{\gamma\in\mathbb{R}^{p_{21}}} \left|\rho_{11}(\hat{\Sigma}_1, \hat{\Sigma}_{21,21}, \mathbb{P}_n, \sqrt{n}(\hat{\beta}_{2,1} - \beta^*_{2,1,n}), \sqrt{n}\beta^*_{2,1,n}, \gamma, \lambda_n)\right| \to_{P_n} 0.$

Proof. First it is easy to verify that $|[H_{2,1}^\intercal\nu - H_{2,1}^\intercal\gamma]_+ - [H_{2,1}^\intercal\gamma]_+| \le |h_{2,1}^\intercal\nu|$. Thus for any probability measure μ in $l^\infty(\widetilde{\mathcal{F}}_{11})$,

$$
\begin{aligned}
|\rho_{11}(\Sigma_1, \Sigma_{21,21}, \mu, \nu, \eta, \gamma, \lambda)| \le K\Bigg\{ &\mu\left(||B_1||\,||H_{2,1}||\,1_{H_{2,1}^\intercal\beta^*_{2,1}=0, \frac{H_{2,1}^\intercal\eta}{||H_{2,1}||} > \sqrt{\lambda k} - K}\right)\\
&+ \mu\left(||B_1||\,||H_{2,1}||\,1_{H_{2,1}^\intercal\beta^*_{2,1}=0, \frac{H_{2,1}^\intercal\eta}{||H_{2,1}||} < -\sqrt{\lambda k} - K}\right)\\
&+ \mu\left(||B_1||\,||H_{2,1}||\,1_{H_{2,1}^\intercal\beta^*_{2,1}\neq 0, -\sqrt{\lambda/k} - K \le \frac{H_{2,1}^\intercal\eta}{||H_{2,1}||} \le \sqrt{\lambda/k} + K}\right)\Bigg\}
\end{aligned}
$$

for a sufficiently large constant $K > 0$ and a sufficiently small constant $k \in (0, 1)$. Since k is held constant there is no loss in generality taking $k = 1$. Define $\rho'_{11} : l^\infty(\mathcal{F}'_{11}) \times \mathbb{R}^{p_{21}} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ as

$$
\begin{aligned}
\rho'_{11}(\mu, \eta, \delta, \delta') = &\mu\left(||B_1||\,||H_{2,1}||\,1_{H_{2,1}^\intercal\beta^*_{2,1}=0, \frac{H_{2,1}^\intercal\eta}{||H_{2,1}||} > \delta}\right)\\
&+ \mu\left(||B_1||\,||H_{2,1}||\,1_{H_{2,1}^\intercal\beta^*_{2,1}=0, \frac{H_{2,1}^\intercal\eta}{||H_{2,1}||} < \delta'}\right)\\
&+ \mu\left(||B_1||\,||H_{2,1}||\,1_{H_{2,1}^\intercal\beta^*_{2,1}\neq 0, \delta' \le \frac{H_{2,1}^\intercal\eta}{||H_{2,1}||} \le -\delta'}\right), \qquad (12)
\end{aligned}
$$

where $\mathcal{F}'_{11} = \Big\{ f(b_1, h_{2,1}) = ||b_1||\,||h_{2,1}||1_{h_{2,1}^\intercal \beta^*_{2,1}=0, \frac{h_{2,1}^\intercal \eta}{||h_{2,1}||} > \delta} + ||b_1||\,|||h_{2,1}||1_{h_{2,1}^\intercal \beta^*_{2,1}=0, \frac{H_{2,1}^\intercal \eta}{||h_{2,1}||} < \delta'} +$ $||b_1||\,||h_{2,1}||1_{h_{2,1}^\intercal \beta^*_{2,1} \neq 0, \delta' \leq \frac{h_{2,1}^\intercal \eta}{||h_{2,1}||} \leq -\delta'}, \eta \in \mathbb{R}^{p_{21}}, \max\{||\eta||, ||\delta||, ||\delta'||\} \leq K \Big\}$. Then

$$|\rho_{11}(\Sigma_1, \Sigma_{21,21}, \mu, \nu, \eta, \gamma, \lambda)| \leq K\rho'_{11}\left(\mu, \eta/\sqrt{n}, (\sqrt{\lambda} - K)/\sqrt{n}, -(\sqrt{\lambda} + K)/\sqrt{n}\right)$$

for $\mu \in l^\infty(\widetilde{\mathcal{F}}_{11})$. In particular for n sufficiently large,

$$\begin{aligned}\Big|\rho_{11}(\hat{\Sigma}_1^{(b)}, \hat{\Sigma}_{21,21}^{(b)}, \hat{\mathbb{P}}_n^{(b)}, \sqrt{n}(\hat{\beta}_{2,1}^{(b)} - \hat{\beta}_{2,1}), \sqrt{n}\hat{\beta}_{2,1}, \gamma, \lambda_n)\Big| \leq \\ K\,\rho'_{11}\left(\hat{\mathbb{P}}_n^{(b)}, \hat{\beta}_{2,1}, (\sqrt{\lambda_n} - K)/\sqrt{n}, -(\sqrt{\lambda_n} - K)/\sqrt{n}\right) \\ + ||c||\,||\hat{\Sigma}_1^{(b)}||\,||\sqrt{n}(\hat{\beta}_{2,1}^{(b)} - \hat{\beta}_{2,1})||\,\hat{\mathbb{P}}_n^{(b)}\left(||B_1||\,||H_{2,1}||\right) 1_{||\sqrt{n}(\hat{\beta}_{2,1}^{(b)} - \hat{\beta}_{2,1})|| > K},\end{aligned}$$

where we have assumed, without loss of generality, that $\hat{\Sigma}_{21,21}^{(b)}$ is the identity matrix. By part 2 of Lemma 1.8 below, we see that the first term on the right hand side of the above display is $o_{P_M}(1)$ almost surely P. To deal with the second term, for any $\epsilon, \delta > 0$, let K sufficiently large so that $P_M\left(\big|\big|\sqrt{n}(\hat{\beta}_{2,1}^{(b)} - \hat{\beta}_{2,1})\big|\big| > K\right) < \delta$ for sufficiently large n for almost all sequences P. Then

$$\begin{aligned}P_M\left(||c||\,||\hat{\Sigma}_1^{(b)}||\,||\sqrt{n}(\hat{\beta}_{2,1}^{(b)} - \hat{\beta}_{2,1})||\hat{\mathbb{P}}_n^{(b)}||B_1||\,||H_{2,1}||1_{|\sqrt{n}(\hat{\beta}_{2,1}^{(b)} - \hat{\beta}_{2,1})| > K} > \epsilon\right) \\ \leq P_M\left(\big|\big|\sqrt{n}(\hat{\beta}_{2,1}^{(b)} - \hat{\beta}_{2,1})\big|\big| > K\right) \leq \delta,\end{aligned}$$

almost surely P. This completes the proof of result 2. Similar arguments can be used to prove results 1 and 3, and are omitted. □

Lemma 1.8. *Let ρ'_{11} be defined in (12). Assume (A1)-(A3), then*

1. $\rho'_{11}(\mathbb{P}_n, \beta^*_{2,1}, (\sqrt{\lambda_n} - K)/\sqrt{n}, (-\sqrt{\lambda_n} - K)/\sqrt{n}) \to_P 0$, *and*

2. $\rho'_{11}(\hat{\mathbb{P}}_n^{(b)}, \hat{\beta}_{2,1}, (\sqrt{\lambda_n} - K)/\sqrt{n}, (-\sqrt{\lambda_n} - K)/\sqrt{n}) \to_{P_M} 0$, *$P$-almost surely.*

If, in addition, we assume (A4), then

3 $\rho'_{11}(P_n, \beta^*_{2,1,n}, (\sqrt{\lambda_n} - K)/\sqrt{n}, (-\sqrt{\lambda_n} - K)/\sqrt{n}) \to_{P_n} 0.$

Proof. The class $\mathcal{F}'_{11}$ is P-Donsker and measurable by Theorem 8.14 in Anthony and Bartlett (2002) and Donkser preservation results (for example, see Theorem 2.10.6 in van der Vaart and Wellner 1996). Note that by (A1) and (A4) $\sup_{f\in\mathcal{F}'_{11}} |Pf^2| < \infty$ and $\sup_{f\in\mathcal{F}'_{11}} |P_n f^2|$ is a bounded sequence. Thus, it follows that (i) $||\mathbb{P}_n - P|| \to 0$ almost surely under P in $l^\infty(\mathcal{F}'_{11})$, (ii) $||\hat{\mathbb{P}}^{(b)}_n - P|| \to 0$ almost surely P_M for almost all sequences P (Lemma 3.6.16 in van der Vaart and Wellner 1996), and (iii) $||\mathbb{P}_n - P_n|| \to 0$ almost surely under P_n in $l^\infty(\mathcal{F}'_{11})$ (Theorem 3.10.12 in van der Vaart and Wellner 1996). Additionally, the argument in the proof of Lemma (1.3) shows that $\hat{\Sigma}_1$ is convergent to Σ_1 under P_n, and the weak law of large numbers establishes convergence under P. The bootstrap strong law shows that $\hat{\Sigma}^{(b)}_1$ converges to Σ_1 in P_M probability for almost all sequences P.

Next we show that ρ'_{11} is continuous at the point $(P, \beta^*_{2,1}, 0, 0)$. Let $\mu_n \to P$ in $l^\infty(\mathcal{F}'_{11})$, $\eta_n \to \beta^*_{2,1}$, $\delta_n \to 0$, and $\delta'_n \to 0$. We have

$$\left|\rho'_{11}(\mu_n, \eta_n, \delta_n, \delta'_n) - \rho'_{11}(P, \beta^*_{2,1}, 0, 0)\right| \le \left|\rho'_{11}(P, \eta_n, \delta_n, \delta'_n) - \rho'_{11}(P, \beta^*_{2,1}, 0, 0)\right| + ||\mu_n - P||,$$

which converges to zero by the dominated convergence theorem. The results follow from the continuous mapping theorems and the fact that $\rho'_{11}(P, \beta^*_{2,1}, 0, 0) = 0$. □

Continuity of w_{11} and w_{12}

To prove Theorems 2.1 and 2.2, we need to show that w_{11} is continuous at points in $(\Sigma_{1,\infty}, \Sigma_{12,\infty}, C_b(\mathcal{F}_{11}), P, \mathbb{R}^{p_2}, (\beta_1^{*\intercal}, \beta_2^{*\intercal})^\intercal)$, $w_{12}(\cdot, \cdot, \cdot, \sqrt{n}\beta^*_{2,1})$ and $w_{12}(\cdot, \cdot, \cdot, \sqrt{n}\beta^*_{2,1,n})$ are continuous at points in $(\Sigma_{1,\infty}, P, \mathbb{R}^{p_{21}})$, and $w'_{12}(\Sigma_1, \mu, \nu) \triangleq \sup_{\gamma\in\mathbb{R}^{p_{21}}} w_{12}(\Sigma_1, \mu, \nu, \gamma)$ is continuous at points in $(\Sigma_{1,\infty}, P, \mathbb{R}^{p_{21}})$.

To prove the desired continuity of w_{12} and w'_{12}, we will establish the stronger result that w_{12} is continuous at points $(\Sigma_{1,\infty}, P, \mathbb{R}^{p_{21}}, \gamma)$ uniformly in γ. That is, for any $\Sigma_n \to \Sigma_{1,\infty}$, probability measures $\mu_n \to P$ and $\nu_n \to \nu$, we have

$$\sup_{\gamma} \Big| w_{12}(\Sigma_n, \mu_n, \nu_n, \gamma) - w_{12}(\Sigma_1, P, \nu, \gamma) \Big| \to 0.$$

Note that

$$\begin{aligned}
&\big| w_{12}(\Sigma_n, \mu_n, \nu_n, \gamma) - w_{12}(\Sigma_1, P, \nu, \gamma) \big| \\
&\quad \le \big| w_{12}(\Sigma_n, \mu_n, \nu_n, \gamma) - w_{12}(\Sigma_n, \mu_n, \nu, \gamma) \big| + \big| w_{12}(\Sigma_n, P, \nu, \gamma) - w_{12}(\Sigma_1, P, \nu, \gamma) \big| \\
&\qquad\qquad + \big| w_{12}(\Sigma_n, \mu_n, \nu, \gamma) - w_{12}(\Sigma_n, P, \nu, \gamma) \big| \\
&\quad \le \mu_n \left(\big| c^\intercal \Sigma_n^{-1} B_1 | H_{2,1}^\intercal (\nu_n - \nu) | \big| \right) + P \left(| c^\intercal (\Sigma_n^{-1} - \Sigma_{1,\infty}^{-1}) B_1 | \, | H_{2,1}^\intercal \nu | \right) \\
&\qquad\qquad + \Big| (\mu_n - P) \Big(c^\intercal \Sigma_n^{-1} B_1 ([H_{2,1}^\intercal \nu + H_{2,1}^\intercal \gamma]_+ - [H_{2,1}^\intercal \gamma]_+) 1_{H_{2,1}^\intercal \beta_{2,1}^* = 0} \Big) \Big|
\end{aligned}$$

By (A2), we have that $||\Sigma_n^{-1}||$ is bounded above for sufficiently large n, where $|| \cdot ||$ of a matrix denotes the spectral norm of the matrix. Thus the first term in the above display is bounded by $||c|| \, ||\Sigma_n^{-1}|| \mu_n(||B_1|| \, ||H_{2,1}||) \, ||\nu_n - \nu|| = o(1)$, and the second term in the above display is bounded by $||c|| \, ||\Sigma_1^{-1} - \Sigma_n^{-1}|| \, P(||B_1|| \, ||H_{2,1}||) ||\nu|| = o(1)$. For the third term, note that if $||\nu|| = 0$, then it is zero. Otherwise,

$$\begin{aligned}
&\Big| (\mu_n - P) \Big(c^\intercal \Sigma_n^{-1} B_1 ([H_{2,1}^\intercal \nu + H_{2,1}^\intercal \gamma]_+ - [H_{2,1}^\intercal \gamma]_+) 1_{H_{2,1}^\intercal \beta_{2,1}^* = 0} \Big) \Big| \\
&\quad \le \Big| (\mu_n - P) \Big(c^\intercal \Sigma_n^{-1} B_1 ([H_{2,1}^\intercal \nu / ||\nu|| + H_{2,1}^\intercal \gamma / ||\nu||]_+ - [H_{2,1}^\intercal \gamma / ||\nu||]_+) 1_{H_{2,1}^\intercal \beta_{2,1}^* = 0} \Big) \Big| \, ||\nu|| \\
&\qquad\qquad \le ||\mu_n - P||_{\mathcal{F}_{12}} ||\nu|| = o(1).
\end{aligned}$$

This established the continuity of w_{12} and hence w'_{12}. The continuity of w_{11} can be established through similar arguments and is therefore omitted.

Remark 1.9 (Plug-in Pretesting Approach)**.** *A natural approach to constructing a confidence interval in a non-regular problem is "a plug-in pretesting approach." This approach, is similar in spirit to the ACI in that it partitions the training data using a series of hypothesis tests and uses different approximations on each partition. In particular, the plug-in pretesting estimator of* $c^\intercal\sqrt{n}(\hat{\beta}_1 - \beta_1^*)$ *is given by*

$$c^\intercal\mathbb{W}_n + c^\intercal\hat{\Sigma}_1^{-1}\mathbb{P}_n B_1^\intercal\mathbb{U}_n 1_{\hat{T}(H_{2,1})>\lambda_n} + c^\intercal\hat{\Sigma}_1^{-1}\mathbb{P}_n B_1^\intercal\left[H_{2,1}^\intercal\mathbb{V}_n\right]_+ 1_{\hat{T}(H_{2,1})\leq\lambda_n}. \tag{13}$$

Confidence intervals are formed by bootstrapping this estimator. Under fixed alternatives, the plug-in pretesting estimator (PPE) is consistent. This consistency is established by recalling that $1_{\hat{T}(h_{2,1})\leq\lambda_n} \to 1_{h_{2,1}^\intercal\beta_{2,1}^*=0}$ *uniformly over sets of probability near one (see the next section).*

However intuitive, the PPE does not perform well in small samples under some generative models (see the main body of the paper and the last section of this supplement). One explanation for this underperformance is that the PPE is not consistent under local alternatives. In particular, under a local generative model as described in (A4), it can be shown that the difference between the PPE and $c^\intercal\sqrt{n}(\hat{\beta}_1 - \beta_{1,n}^*)$ *is equal to*

$$c^\intercal\hat{\Sigma}_1^{-1}\mathbb{P}_n B_1^\intercal\left(\left[H_{2,1}^\intercal\left(\mathbb{V}_n^{(n)}+\gamma\right)\right]_+ - \left[H_{2,1}^\intercal\gamma\right]_+ - \left[H_{2,1}^\intercal\mathbb{V}_n^{(n)}\right]_+\right)1_{H_{2,1}^\intercal\beta_{2,1}^*=0} + o_{P_n}(1), \tag{14}$$

which is does not vanish for any alternative γ *for which* $H_{2,1}^\intercal\gamma$ *is not identically zero with probability one.*

The expression in (14) offers yet another view of the ACI. In particular, one can view the last term of $\mathcal{U}(c)$ *as approximating the supremum over local alternatives of the difference between the PPE and the target* $c^\intercal\sqrt{n}(\hat{\beta}_1 - \beta_{1,n}^*)$*. In this way, the ACI can be thought of as a corrected version of the PPE where the correction is intended to safeguard against poor small sample performance.*

1.3 Note on computation

The time complexity of computing the ACI bounds for a bootstrap sample is not much larger than the $O(n^3)$ time required to compute the QR decomposition of the design matrix for least-squares linear regression. The most computationally expensive addition to the standard bootstrap is the approximation of the supremum in the third term of equation (9) in the main body of the paper (here after (11)), and its corresponding infimum for the lower bound.

In order to approximate the supremum in (11), we use a simple stochastic search. Since the intuition is that we want to take the supremum over a region near $\sqrt{n}\hat{\beta}_{2,1}$, we draw n_γ candidate vectors $\gamma^1, \gamma^2, ..., \gamma^{n_\gamma}$ uniformly from within a large set centered at $\hat{\beta}_{2,1}$. We then evaluate the supremum objective in the third term of (11) at each of the γ^i. Similarly, we use the minimum to approximate the infimum in the lower bound. In our experiments, we used $n_\gamma = 1000$. The complexity of evaluating the supremum objective once is $n \times K_t$, so the additional time added to the standard bootstrap procedure is $O(n_\gamma \cdot n \cdot K_t)$.

2 Extension of the ACI to more than two stages and more than two treatments

In this appendix, we develop the ACI for the general case where there is an arbitrary finite number of stages of treatment, and an arbitrary finite number of treatment choices at each stage. We begin with a review of the Q-learning procedure in this setting.

2.1 Q-Learning in the general case

We observe an *i.i.d.* sample of trajectories $\{\mathcal{T}_i\}_{i=1}^n$ drawn from a fixed but unknown distribution P. Each trajectory is of the form

$$\mathcal{T} = (X_1, A_1, Y_1, X_2, A_2, Y_2, \ldots, X_T, A_T, Y_T), \tag{15}$$

being comprised of patient measurements X_t, assigned treatment A_t, and observed response Y_t for $t = 1, 2, \ldots, T$. For each decision point t the assigned treatment A_t is coded to take values in the set $\{1, 2, \ldots, K_t\}$. As in the two-stage setting, we let H_t denote a concise summary of patient history at time t. More precisely, $H_1 \triangleq \Psi_1(X_1)$ and $H_t \triangleq \Psi_t(X_1, A_1, Y_1, \ldots, X_{t-1}, A_{t-1}, Y_{t-1}, X_t)$ for $t = 2, 3, \ldots, T$ for known functions Ψ_t. The form of the working model for the Q-function is of the same form as in Section 3 of the main body of the paper. For each t we use the model

$$Q_t(h_t, a_t; \beta_t) \triangleq \sum_{i=1}^{K_t} h_{t,1}^{\intercal} 1_{a_t=i} \beta_{t,i}, \tag{16}$$

where $\beta_t \triangleq (\beta_{t,1}^{\intercal}, \beta_{t,2}^{\intercal}, \ldots, \beta_{t,K_t}^{\intercal})^{\intercal} \in \mathbb{R}^{p_t}$. The form of the foregoing model is to produce compact theoretical expressions. In practice, any coding that makes the model identifiable can be used. Note, however, that the form of the pretest will depend on the coding. As in the two stage setting, the form of the working model implies that when $h_{t,1}^{\intercal}\beta_{t,i} - \max_{j\neq i} h_{t,1}^{\intercal}\beta_{t,j} \approx 0$ for some $1 \leq i \leq K_t$, then at least two treatments are approximately optimal for a patient with history $H_{t,1} = h_{t,1}$. That is, there is *not* a unique best treatment for a patient with history $H_{t,1} = h_{t,1}$. On the other hand, if $\left|h_{t,1}^{\intercal}\beta_{t,i} - \max_{j\neq i} h_{t,1}^{\intercal}\beta_{t,j}\right| \gg 0$ for all $1 \leq i \leq K_t$, then the working model implies that *exactly* one treatment is best for a patient with history $H_{t,1} = h_{t,1}$. Once a working model has been specified, the Q-learning algorithm can be applied to estimate the optimal DTR. The Q-learning algorithm is a follows:

1. Regress Y_T on H_T and A_T using the working model (16) to obtain:

$$\hat{\beta}_T \triangleq \arg\min_{\beta_T} \mathbb{P}_n \left(Y_T - Q_T(H_T, A_T; \beta_T)\right)^2$$

   and subsequently the approximation $Q_T(h_T, a_T; \hat{\beta}_T)$ to the conditional mean $Q_T(h_T, a_T)$.

2. (a) Recursively, define the predicted future reward following the optimal policy as:

$$\begin{aligned} \tilde{Y}_t &\triangleq Y_t + \max_{a_{t+1} \in \{1,2,\ldots,K_{t+1}\}} Q_{t+1}\left(H_{t+1}, a_{t+1}; \hat{\beta}_{t+1}\right) \\ &= Y_t + \max_{1 \leq i \leq K_{t+1}} H_{t+1,1}^{\intercal} \hat{\beta}_{t+1,i} \end{aligned}$$

   for $t = T-1, T-2, \ldots, 1$.

   (b) Regress $\tilde{Y}_t$ on H_t and A_t using the working model (16) to obtain $\hat{\beta}_t \triangleq \arg\min_{\beta_t} \mathbb{P}_n \left(\tilde{Y}_t - Q_t(H_t, A_t; \beta_t)\right)^2$.

3. Define the estimated optimal DTR $\hat{\pi} = (\hat{\pi}_1, \hat{\pi}_2, \ldots, \hat{\pi}_T)$ so that

$$\hat{\pi}_t(h_t) \triangleq \arg \max_{a_t \in \{1,2,\ldots,K_t\}} Q_t(h_t, a_t; \hat{\beta}_t).$$

When $T = 2$ the above procedure is equivalent to the two stage Q-learning algorithm given in Section 3 of the main body of the paper.

Our aim is to use the ACI to construct a confidence interval for $c^{\intercal}\beta_1^*$ where c is an arbitrary vector of constants. The definition of β_1^* is given inductively. Define

$$\beta_T^* \triangleq \arg\min_{\beta_T} P\left(Y_T - Q_T(H_T, A_T; \beta_T)\right)^2.$$

For $t = T-1, T-2, \ldots, 1$ define

$$\begin{aligned}
\tilde{Y}_t^* &\triangleq Y_t + \max_{1 \leq i \leq K_{t+1}} H_{t+1,1}^{\intercal} \beta_{t+1,i}^*, \\
\beta_t^* &\triangleq \arg\min_{\beta_t} P \left(\tilde{Y}_T^* - Q_t(H_t, A_t; \beta_t) \right)^2 .
\end{aligned}$$

At times we will focus on the problem of constructing a confidence interval for a linear combination of the first stage coefficients β_1^* since building a confidence interval for, say $c_t^{\intercal} \beta_t^*$, is equivalent to building a confidence interval for the first stage of a $T-t+1$ stage trial. That is, one can always view the tth stage as the first stage of a shorter $T-t+1$ stage trial. Information collected prior to the tth stage can be treated as baseline (pre-randomization) information in this conceptual shorter trial.

2.2 ACI in the general case

The ACI in the general case is conceptually the same as the two stage case. Non-regularity in $\sqrt{n}(\hat{\beta}_t - \beta_t^*)$ arises whenever there are two or more equally best treatments at *any* future stage of treatment $s > t$ for a non-null subset of patient histories. The ACI works by constructing smooth upper and lower bounds on $\sqrt{n}(\hat{\beta}_t - \beta_t^*)$ and then bootstrapping these bounds to construct confidence intervals. As in the two stage case, these bounds are asymptotically equivalent to taking the supremum (infimum) over all local alternatives to the true generative distribution.

In order to develop the ACI in this general setting, we generalize the notation given in the main body of the paper. Define $B_t \triangleq (H_{t,1}^{\intercal} 1_{A_t=1}, \ldots, H_{t,1}^{\intercal} 1_{A_t=K_t})^{\intercal}$ so that instances of B_t form the columns of the design matrix used in the tth stage regression. Further, define $\hat{\Sigma}_t \triangleq \mathbb{P}_n B_t B_t^{\intercal}$. The limiting distribution of $\sqrt{n}(\hat{\beta}_t - \beta_t^*)$ depends abruptly on the frequency of patients for which there are multiple equally optimal best treatments at a future stage.

Consequently, the set

$$\mathcal{A}_t^*(h_{t,1}) \triangleq \arg\max_{1\le i\le K_t} h_{t,1}^T \beta_{t,i}^* \tag{17}$$

of equally optimal treatments at stage t for a patient with history $H_{t,1} = h_{t,1}$, is relevant for the development of asymptotic theory. Notice that $\mathcal{A}_t^*(h_{t,1})$ is a singleton when there is exactly one best treatment for a patient with history $h_{t,1}$. As in the two stage case, we will need an estimator of $\mathcal{A}_t^*(h_{t,1})$. The estimator we use is based on the following test statistics:

$$\hat{T}_{t,i}(h_{t,1}) \triangleq \frac{n\left(h_{t,1}^\intercal \hat{\beta}_{t,i} - \max_{j\neq i} h_{t,1}^\intercal \hat{\beta}_{t,j}\right)^2}{h_{t,1}^\intercal \hat{\zeta}_{t,i} h_{t,1}} \tag{18}$$

where $\hat{\zeta}_{t,i}$ is the usual plug-in estimator of $n\mathrm{Cov}(\hat{\beta}_{t,i} - \hat{\beta}_{t,k})$ evaluated at $k = \hat{k}_i$, where $\hat{k}_i = \arg\max_{j\neq i} h_{t,1}^\intercal \hat{\beta}_{t,j}$ (we are acting as if the maximizing index were known a priori. In particular, if $\hat{\sigma}_t^2$ is the empirical mean square error of the tth stage regression, then we use $\hat{\sigma}_t \mathbb{P}_n B_t B_t^\intercal$ as our estimate of the asymptotic covariance of $\hat{\beta}_t$. Thus, $\hat{\zeta}_{t,i}$ is given by $\hat{\Sigma}_{t,ii} - 2\hat{\Sigma}_{t,i\hat{k}_i} + \hat{\Sigma}_{t,\hat{k}_i\hat{k}_i}$ where $\hat{\Sigma}_{t,jl}$ is the submatrix of $\hat{\sigma}_t \mathbb{P}_n B_t B_t^T$ corresponding to the estimator of $Cov(\hat{\beta}_{t,j}, \hat{\beta}_{t,l})$.) The statistic, $\min_i \hat{T}_{t,i}(h_{t,1})$, should be large when there is exactly one best treatment for a patient with history $H_{t,1} = h_{t,1}$. On the other hand, $\min_i \hat{T}_{t,i}(h_{t,1})$ should be small if there is more than one best treatment. Thus, a natural estimator of $\mathcal{A}_t^*(h_{t,1})$ is

$$\hat{\mathcal{A}}_t(h_{t,1}) = \begin{cases} \left\{i \,:\, \hat{T}_{t,i}(h_{t,1}) \le \lambda_n\right\} & \text{if } \min_i \hat{T}_{t,i}(h_{t,1}) \le \lambda_n \\ \arg\max_{1\le i\le K_t} h_{t,1}^\intercal \hat{\beta}_{t,i} & \text{if } \min_i \hat{T}_{t,i}(h_{t,1}) > \lambda_n. \end{cases}$$

The merits and genesis of this statistic were discussed in the main body of the paper. Under the regularity conditions given in the next section, it follows that $\hat{\mathcal{A}}_t(h_{t,1})$ is a consistent estimator of $\mathcal{A}_t^*(h_{t,1})$. In a slight abuse of notation, define $\widetilde{\mathcal{A}}_t(h_{t,1}) \triangleq \hat{\mathcal{A}}_t(h_{t,1}) \bigcup \mathcal{A}_t^*(h_{t,1})$ be the union of the estimator and the estimand.

Recall that we denote $\mathbb{V}_{t,n} \triangleq \sqrt{n}(\hat{\beta}_t - \beta_t^*)$ and $\mathbb{V}_{t,n,i} \triangleq \sqrt{n}(\hat{\beta}_{t,i} - \beta_{t,i}^*)$ for $t = 1, \ldots, T$.

For any $\gamma_{t+1} \in \mathbb{R}^{p_{t+1}}$ and $\Gamma_{t+1} = (\gamma_{t+1}^\intercal, \gamma_{t+2}^\intercal, \ldots, \gamma_T^\intercal)^\intercal \in \mathbb{R}^{\sum_{k=t+1}^T p_k}$, $t = 1, \ldots, T-1$, define

$$\begin{aligned}\widetilde{\mathbb{V}}_{T-1,n}(\Gamma_T) &\triangleq \mathbb{W}'_{T-1,n} + \hat{\Sigma}_{T-1}^{-1}\mathbb{P}_n B_{T-1}\mathbb{U}_{T,n} 1_{\#\hat{\mathcal{A}}_T(H_{T,1})=1} \\ &\quad + \hat{\Sigma}_{T-1}^{-1}\mathbb{P}_n B_{T-1}\left[\max_{i\in\widetilde{\mathcal{A}}_T(H_{T,1})} H_{T,1}^\intercal(\mathbb{V}_{T,n,i} + \gamma_{T,i}) - \max_{i\in\widetilde{\mathcal{A}}_T(H_{T,1})} H_{T,1}^\intercal\gamma_{T,i}\right] 1_{\#\hat{\mathcal{A}}_T(H_{T,1})>1},\end{aligned}$$

and for $t < T-1$,

$$\begin{aligned}\widetilde{\mathbb{V}}_{t,n}(\Gamma_{t+1}) &\triangleq \mathbb{W}'_{t,n} + \hat{\Sigma}_t^{-1}\mathbb{P}_n B_t \max_{i\in\hat{\mathcal{A}}_{t+1}(H_{t+1,1})} H_{t+1,1}^\intercal \widetilde{\mathbb{V}}_{t+1,n,i}(\Gamma_{t+2}) 1_{\#\hat{\mathcal{A}}_{t+1}(H_{t+1,1})=1} \\ &\quad + \hat{\Sigma}_t^{-1}\mathbb{P}_n B_t\left(\mathbb{U}_{t+1,n} - \max_{i\in\hat{\mathcal{A}}_{t+1}(H_{t+1,1})} H_{t+1,1}^\intercal \mathbb{V}_{t+1,n,i}\right) 1_{\#\hat{\mathcal{A}}_{t+1}(H_{t+1,1})=1} \\ &+ \hat{\Sigma}_t^{-1}\mathbb{P}_n B_t\left(\max_{i\in\widetilde{\mathcal{A}}_{t+1}(H_{t+1,1})} H_{t+1,1}^\intercal\left(\widetilde{\mathbb{V}}_{t+1,n,i}(\Gamma_{t+2}) + \gamma_{t+1,i}\right) - \max_{i\in\widetilde{\mathcal{A}}_{t+1}(H_{t+1,1})} H_{t+1,1}^\intercal\gamma_{t+1,i}\right) \\ &\qquad \times 1_{\#\hat{\mathcal{A}}_{t+1}(H_{t+1,1})>1},\end{aligned}$$

where

$$\begin{aligned}\mathbb{W}'_{t,n} &= \hat{\Sigma}_t^{-1}\sqrt{n}\mathbb{P}_n B_t\left\{Y_t + \max_{i=1,\ldots,K_{t+1}} H_{t+1,1}^\intercal\beta_{t+1,i}^* - B_t\beta_t^*\right\} \\ \mathbb{U}_{t,n} &= \sqrt{n}\left(\max_{i=1,\ldots,K_{t+1}} H_{t,1}^\intercal\hat{\beta}_{t,i} - \max_{i=1,\ldots,K_{t+1}} H_{t,1}^\intercal\beta_{t,i}^*\right).\end{aligned}$$

for $t = 1, 2, \ldots, T-1$. Then $\widetilde{\mathbb{V}}_{t,n}((\sqrt{n}\beta_{t+1}^{*\intercal}, \ldots, \sqrt{n}\beta_T^{*\intercal})^\intercal) = \sqrt{n}(\hat{\beta}_t - \beta_t^*)$ for $t = 1, \ldots, T-1$. The upper bound on $c_t^\intercal\sqrt{n}(\hat{\beta}_t - \beta_t^*)$ used to construct a confidence interval for $c_t^\intercal\beta_t^*$ is given by $\mathcal{U}_t(c_t) \triangleq \sup_{\Gamma_{t+1}\in\mathbb{R}^{\sum_{k=t+1}^T p_k}} c_t^\intercal\widetilde{\mathbb{V}}_{t,n}(\Gamma_{t+1})$. Similarly, the lower bound is obtained by replacing the sup with an inf.

2.2.1 Example: ACI for three stages

To illustrate the ACI for the general case and solidify the ideas presented in the preceding section, we provide the bounds for the case where there are three stages of treatment and an arbitrary number of treatments at each stage. Thus, $T = 3$ and $\mathbb{V}_{3,n} \triangleq \sqrt{n}(\hat{\beta}_3 - \beta_3^*)$. Since $\mathbb{V}_{3,n}$ is the usual least squares estimator, it follows under (A1)-(A2) (see below) that $\mathbb{V}_{3,n}$ is regular and its limiting distribution is normal. The process $\widetilde{\mathbb{V}}_{2,n}(\gamma_3)$ indexed by $\gamma_3 \in \mathbb{R}^{p_3}$ is defined as follows

$$\begin{aligned}\widetilde{\mathbb{V}}_{2,n}(\gamma_3) \triangleq \mathbb{W}'_{2,n} + \hat{\Sigma}_2^{-1}\mathbb{P}_n B_2 \mathbb{U}_{3,n} 1_{\#\hat{\mathcal{A}}_3(H_{3,1})=1} \\ + \hat{\Sigma}_2^{-1}\mathbb{P}_n B_2 \left[\max_{i \in \widetilde{\mathcal{A}}_3(H_{3,1})} H_{3,1}^{\intercal} \left(\mathbb{V}_{3,n,i} + \gamma_{3,i}\right) - \max_{i \in \widetilde{\mathcal{A}}_3(H_{3,1})} H_{3,1}^{\intercal}\gamma_{3,i}\right] 1_{\#\hat{\mathcal{A}}_3(H_{3,1})>1}.\end{aligned}$$

An ACI for $c_2^{\intercal}\beta_2^*$ is formed using the bootstrap distribution of bounds $\mathcal{U}_2(c_2) \triangleq \sup_{\gamma_3 \in \mathbb{R}^{p_3}} c_2^{\intercal}\widetilde{\mathbb{V}}_{2,n}(\gamma_3)$ and $\mathcal{L}_2(c_2) \triangleq \inf_{\gamma_3 \in \mathbb{R}^{p_3}} c_2^{\intercal}\widetilde{\mathbb{V}}_{2,n}(\gamma_3)$.

To form a confidence interval for the first stage coefficients, e.g. $c_1^{\intercal}\beta_1^*$, we use the process $\widetilde{\mathbb{V}}_{1,n}((\gamma_2, \gamma_3))$ indexed by $(\gamma_2, \gamma_3) \in \mathbb{R}^{p_2+p_3}$. The definition of $\widetilde{\mathbb{V}}_{1,n}((\gamma_2, \gamma_3))$ is given by

$$\begin{aligned}\widetilde{\mathbb{V}}_{1,n}((\gamma_2, \gamma_3)) \triangleq \mathbb{W}'_{1,n} + \hat{\Sigma}_1^{-1}\mathbb{P}_n B_1 \max_{i \in \hat{\mathcal{A}}_2(H_{2,1})} H_{2,1}^{\intercal}\widetilde{\mathbb{V}}_{2,n,i}(\gamma_3) 1_{\#\hat{\mathcal{A}}_2(H_{2,1})=1} \\ + \hat{\Sigma}_1^{-1}\mathbb{P}_n B_1 \left(\mathbb{U}_{2,n} - \max_{i \in \hat{\mathcal{A}}_2(H_{2,1})} H_{2,1}^{\intercal}\mathbb{V}_{2,n,i}\right) 1_{\#\hat{\mathcal{A}}_{2,1}(H_{2,1})=1} \\ + \hat{\Sigma}_1^{-1}\mathbb{P}_n B_1 \left(\max_{i \in \widetilde{\mathcal{A}}_2(H_{2,1})} H_{2,1}^{\intercal}\left(\widetilde{\mathbb{V}}_{2,n,i}(\gamma_3) + \gamma_{2,i}\right) - \max_{i \in \widetilde{\mathcal{A}}_2(H_{2,1})} H_{2,1}^{\intercal}\gamma_{2,i}\right) 1_{\#\hat{\mathcal{A}}_2(H_{2,1})>1}.\end{aligned}$$

Thus, the upper and lower bounds used for constructing a confidence interval for $c_1^{\intercal}\beta_1^*$ are given by $\mathcal{U}_1(c_1) \triangleq \sup_{\gamma_2 \in \mathbb{R}^{p_2}, \gamma_3 \in \mathbb{R}^{p_3}} c_1^{\intercal}\widetilde{\mathbb{V}}_{1,n}((\gamma_2, \gamma_3))$ and $\mathcal{L}_1(c_1) \triangleq \inf_{\gamma_2 \in \mathbb{R}^{p_2}, \gamma_3 \in \mathbb{R}^{p_3}} c_1^{\intercal}\widetilde{\mathbb{V}}_{1,n}((\gamma_2, \gamma_3))$. The form of $\widetilde{\mathbb{V}}_{2,n}(\gamma_3)$ and $\widetilde{\mathbb{V}}_{1,n}((\gamma_2, \gamma_3))$ show that computing the bounds $\mathcal{U}_1(c_1)$ and $\mathcal{L}_1(c_1)$ require optimizing piecewise linear objective functions. Since these piecewise linear objec-

tives are non-convex (non-concave) the resultant optimization problem is, to the best of our knowledge, a mixed integer program. A simple stochastic approximation is given in Section 1.4 of this supplement.

2.3 Properties of the ACI in the general case

In this section, we state the general case analogs of the theorems given in the main body of the paper. In particular, these results state that the ACI provides asymptotically valid confidence intervals under mild regularity conditions. In addition, under further assumptions, it can be shown that the ACI delivers asymptotically exact coverage.

We will make the following assumptions.

(A1) The histories H_t, features B_t, and outcomes Y_t, satisfy the moment inequalities $P||H_t||^2\,||B_{t-1}||^2 < \infty$ for all $t = 2, 3, \ldots, T$, and $PY_t^2\,||B_t||^2 < \infty$ for all $t = 1, 2, \ldots, T$.

(A2) Define:

1. $\Sigma_{t,\infty} \triangleq PB_tB_t^\intercal$;
2. $g_T(B_T, Y_T; \beta_T^*) \triangleq B_T(Y_T - B_T^\intercal \beta_T^*)$;
3. $g_t(B_t, Y_t, H_{t+1}; \beta_t^*) \triangleq B_t\left(Y_t + \max_{k\in\mathcal{A}_{t+1}^*} H_{t+1,1}^\intercal \beta_{t+1,k}^* - B_t^\intercal \beta_t^*\right)$;

then the matrices $\Sigma_{t,\infty}$ for $t = 1, \ldots, T-1$, and $\Omega \triangleq \mathrm{Cov}\left((g_1^\intercal, g_2^\intercal, \ldots, g_T^\intercal)^\intercal\right)$ are strictly positive definite.

(A3) The sequence λ_n tends to infinity and satisfies $\lambda_n = o(n)$.

(A4) There exists a sequence of local alternatives P_n converging to P in the sense that

$$\int \left[\sqrt{n}(dP_n^{1/2} - dP^{1/2}) - \frac{1}{2} g dP^{1/2}\right]^2 \to 0$$

for some measurable real-valued function g for which $P_n||H_t||^2\,||B_{t-1}||^2$ is a bounded sequence for $t = 2, 3, \dots, T$, and $P_n Y_t^2\,||B_t||^2$ is a bounded sequence for $t = 1, 2, \dots, T$.

These assumptions are quite mild requiring the kind of moment and collinearity constraints which are often encountered in multiple regression. Assumption (A3) concerns a user-controlled parameter and is thus readily satisfied. Assumption (A4) implies that

$$\begin{aligned}
\beta^*_{T,n} &\triangleq \arg\min_{\beta\in\mathbb{R}^{p_T}} P_n(Y_T - B_T^\intercal\beta)^2,\\
\beta^*_{t,n} &\triangleq \arg\min_{\beta\in\mathbb{R}^{p_t}} P_n(Y_t + \max_{i=1,\dots,K_{t+1}} H_{t+1,1}^\intercal\beta_{t+1,n,i} - B_t^\intercal\beta)^2, \quad \text{for} \quad t = 1, 2, \dots, T-1,
\end{aligned}$$

satisfies $\lim_{n\to\infty}\sqrt{n}(\beta^*_{t,n} - \beta^*_t) = \gamma^*_t$ for some $\gamma^*_t \in \mathbb{R}^{p_t}$, $t = 1, \dots, T$.

Let $\widetilde{\mathbb{V}}_{t,\infty}(\Gamma_{t+1})$ denote the limiting process of $\widetilde{\mathbb{V}}_{t,n}(\Gamma_{t+1})$ indexed by $\Gamma_{t+1} \in \mathbb{R}^{\sum_{k=t+1}^T p_k}$, $\mathbb{W}'_{t,\infty}$ denote the limiting distribution of $\mathbb{W}'_{t,n}$, and $\mathbb{V}_{t,\infty}$ denote the limiting distribution of $\mathbb{V}_{t,n} \triangleq \sqrt{n}(\hat\beta_t - \beta^*_t)$. Let $\mathbb{V}_{t,\gamma^*,\infty}$ denote the limiting distribution of $\sqrt{n}(\hat\beta_t - \beta^*_{t,n})$ for $t = 1, \dots, T$. In particular, note that $\mathbb{V}_{T,\gamma^*,\infty} = \mathbb{V}_{T,\infty}$, which does not depend on γ^*_T. Under local alternatives P_n, the analog of $\widetilde{\mathcal{A}}_t(h_{t,1})$ is defined as $\widetilde{\mathcal{A}}_{t,n}(h_{t,1}) \triangleq \hat{\mathcal{A}}_t(h_{t,1}) \bigcup \mathcal{A}^*_{t,n}(h_{t,1})$, where

$$\mathcal{A}^*_{t,n}(h_{t,1}) = \begin{cases} \{i : T_{t,n,i}(h_{t,1}) \le \lambda_n\} & \text{if } \min_i T_{t,n,i}(h_{t,1}) \le \lambda_n \\ \arg\max_{i=1,\dots,K_t} h_{t,1}^\intercal \beta^*_{t,n,i} & \text{if } \min_i T_{t,n,i}(h_{t,1}) > \lambda_n \end{cases},$$

$T_{t,n,i}(h_{t,1}) \triangleq n(h_{t,1}^T\beta^*_{t,n,i} - \max_{j\ne i} h_{t,1}^T\beta^*_{t,n,j})^2/(h_{t,1}^\intercal\zeta_{t,i,\infty}h_{t,1})$ and $\zeta_{t,i,\infty}$ is the limit of $\hat\zeta_{t,i}$ for $i = 1, \dots, K_t$. We have the following theorem.

Theorem 2.1 (Validity of Population Bounds)**.** *Assume (A1)-(A3) and fix $c_t \in \mathbb{R}^{\dim(\beta^*_t)}$.*

1. *The limiting distribution $c_t^\intercal\sqrt{n}(\hat\beta_t - \beta^*_t)$ is given by the distribution of*

$$c_t^\intercal\mathbb{V}_{t,\infty} = c_t^\intercal\mathbb{W}'_{t,\infty} + c_t^\intercal\Sigma^{-1}_{t,\infty}PB_t \max_{i\in\mathcal{A}^*_{t+1}(H_{t+1,1})} H_{t+1,1}^\intercal\mathbb{V}_{t+1,\infty,i}$$

for $t = T-1, \dots, 1$.

2. *If, in addition (A4) holds, then the limiting distribution* $c_t^\intercal \sqrt{n}(\hat{\beta}_t - \beta^*_{t,n})$ *is given by the distribution of*

$$c_t^\intercal \mathbb{V}_{t,\gamma^*,\infty} = c_t^\intercal \mathbb{W}'_{t,\infty} + c_t^\intercal \Sigma_{t,\infty}^{-1} P B_t \max_{i \in \mathcal{A}^*_{t+1}(H_{t+1,1})} H_{t+1,1}^\intercal \mathbb{V}_{t+1,\gamma^*,\infty,i} 1_{\#\mathcal{A}^*_{t+1}(H_{t+1,1})=1}$$
$$+ c_t^\intercal \Sigma_{t,\infty}^{-1} P B_t \left[\max_{i \in \mathcal{A}^*_{t+1}(H_{t+1,1})} H_{t+1,1}^\intercal (\mathbb{V}_{t+1,\gamma^*,\infty,i} + \gamma^*_{t+1,i}) - \max_{i \in \mathcal{A}^*_{t+1}(H_{t+1,1})} H_{t+1,1} \gamma^*_{t+1,i} \right] 1_{\#\mathcal{A}^*_{t+1}(H_{t+1,1})>1}$$

for $t = T-1, \ldots, 1$.

3. *The limiting distribution* $\mathcal{U}_{T-1}(c_{T-1})$ *under* P *or* P_n *is given by the distribution of*

$$c_{T-1}^\intercal \mathbb{W}'_{T-1,\infty} + c_{T-1}^\intercal \Sigma_{T-1,\infty}^{-1} P B_{T-1} \max_{i \in \mathcal{A}^*_T(H_{T,1})} H_{T,1}^\intercal \mathbb{V}_{T,\infty,i} 1_{\#\mathcal{A}^*_T(H_{T,1})=1}$$
$$+ \sup_{\gamma_T \in \mathbb{R}^{p_T}} c_{T-1}^\intercal \Sigma_{T-1,\infty}^{-1} P B_{T-1} \left(\max_{i \in \mathcal{A}^*_T(H_{T,1})} H_{T,1}^\intercal (\mathbb{V}_{T,\infty,i} + \gamma_{T,i}) - \max_{i \in \mathcal{A}^*_T(H_{T,1})} H_{T,1}^\intercal \gamma_{T,i} \right) 1_{\#\mathcal{A}^*_T(H_{T,1})>1};$$

and for $t < T-1$, *the limiting distribution of* $\mathcal{U}_t(c_t)$ *under* P *or* P_n *is given (recursively) by the distribution of*

$$c_t^\intercal \mathbb{W}'_{t,\infty} + \sup_{\Gamma_{t+1} \in \mathbb{R}^{\sum_{k=t+1}^T p_k}} \Bigg\{ c_t^\intercal \Sigma_{t,\infty}^{-1} P B_t \max_{i \in \mathcal{A}^*_{t+1}(H_{t,1})} H_{t+1,1}^\intercal \widetilde{\mathbb{V}}_{t+1,\infty,i}(\Gamma_{t+2}) 1_{\#\mathcal{A}^*_{t+1}(H_{t,1})=1}$$
$$+ c_t^\intercal \Sigma_{t,\infty}^{-1} P B_t \Big[\max_{i \in \mathcal{A}^*_T(H_{T,1})} H_{t+1,1}^\intercal \left(\widetilde{\mathbb{V}}_{t+1,\infty,i}(\Gamma_{t+2}) + \gamma_{t+1,i} \right)$$
$$- \max_{i \in \mathcal{A}^*_T(H_{T,1})} H_{t+1,1}^\intercal \gamma_{t+1,i} \Big] 1_{\#\mathcal{A}^*_{t+1}(H_{t,1})>1} \Bigg\}.$$

When $T = 2$, these limiting distributions are equal in law to the limiting distributions of $\mathcal{U}(c)$ and $\mathcal{L}(c)$ given in Section 2 of the main body of the paper (after appropriate recoding). The preceding theorem shows that the limiting distribution of $\mathcal{U}_t(c_t)$ is stochastically larger than that of $c_t^\intercal \sqrt{n}(\hat{\beta}_t - \beta^*_t)$. A similar result can be stated in terms of a lower bound $\mathcal{L}_t(c_t)$ by replacing the sup by an inf in the preceding theorem. The theorem is proved recursively

using the results proved for the two stage case and then repeatedly invoking the continuous mapping theorem.

In order to form a confidence interval, the bootstrap distributions of $\mathcal{U}_t(c_t)$ and $\mathcal{L}_t(c_t)$, which we denote by $\mathcal{U}_t^{(b)}(c_t)$ and $\mathcal{L}_t^{(b)}(c_t)$, are used. The next result states that the bootstrap bounds are asymptotically consistent.

Theorem 2.2. *Assume (A1)-(A3) and fix* $c_t \in \mathbb{R}^{\dim(\beta_t^*)}$ *for* $t = 1, \ldots, T-1$. *Then* $(\mathcal{U}_t(c_t), \mathcal{L}_t(c_t))$ *and* $(\mathcal{U}_t^{(b)}(c_t), \mathcal{L}_t^{(b)}(c_t))$ *converge to the same limiting distribution in probability for all* t. *That is,*

$$\sup_{\nu \in BL_1(\mathbb{R}^2)} \left| \mathbb{E}\nu\left((\mathcal{U}_t(c_t), \mathcal{L}_t(c_t))\right) - \mathbb{E}_M \nu\left(\left(\mathcal{U}_t^{(b)}(c_t), \mathcal{L}_t^{(b)}(c_t)\right)\right) \right| \to_{P^*} 0$$

for $t = 1, \ldots, T-1$.

Corollary 2.3. *Assume (A1)-(A3) and fix* $c_t \in \mathbb{R}^{\dim(\beta_t^*)}$ *for* $t = 1, \ldots, T-1$. *Let* $\hat{u}$ *denote the* $1-\alpha/2$ *quantile of* $\mathcal{U}_t^{(b)}(c_t)$ *and* $\hat{l}$ *denote the* $\alpha/2$ *quantile of* $\mathcal{L}_t^{(b)}(c_t)$. *Then for all* t

$$P_M\left(c_t^\intercal \hat{\beta}_t - \hat{u}/\sqrt{n} \leq c_t^\intercal \beta_t^* \leq c_t^\intercal \hat{\beta}_t - \hat{l}/\sqrt{n}\right) \geq 1 - \alpha + o_P(1).$$

Furthermore, for a given value of t *if* $P\left(\min_{i=1,\ldots,K_s}\left|H_{s,1}^\intercal \beta_{s,i}^* - \max_{j \neq i} H_{s,1}^\intercal \beta_{s,j}^*\right| = 0\right) = 0$ *for all* $s = t+1, t+2, \ldots, T$, *then the above inequality can be strengthened to equality.*

The preceding result shows that the ACI can be used to construct valid confidence intervals regardless of the underlying parameters or generative model. In addition, when there is almost always a unique best treatment, then the ACI delivers asymptotically exact confidence intervals.

2.4 Sketched proofs for the ACI in three stages and more than two treatments

In this section, we provide proof Sketches for Theorems 2.1 and 2.2 for $T = 3$. Results for $T > 3$ can be proved in a similar fashion and are omitted. Note that $\hat{\zeta}_{t,i}$ used in (18) is for normalization purpose and its value does not affect the asymptotic results. Although by definition $\hat{\zeta}_{t,i}$ depends on $h_{t,1}$, this dependence is only through $\hat{k}_i$ and thus takes a finite number (at most $K_t - 1$) of values. Thus throughout this section we assume without loss of generality that $\hat{\zeta}_{t,i}$ does not depend on $h_{t,1}$ and is positive definite for all n.

Stage 3. For any $c_3 \in \mathbb{R}^{p_3}$, define the function $w_3 : \mathcal{D}_{p_3} \times l^{\infty}(\mathcal{F}_3) \times \mathbb{R}^{p_3} \to \mathbb{R}$ as

$$w_3(\Sigma, \mu, \beta_3) \triangleq \mu\left(c_3^{\intercal}\Sigma^{-1}B_3(Y_3 - B_3^{\intercal}\beta_3)\right),$$

where $\mathcal{F}_3 = \{f(b_3, y_3) = a^{\intercal}b_3(y_3 - b_3^{\intercal}\beta_3) \,:\, a, \beta_3 \in \mathbb{R}^{p_3},\ \max\{||a||, ||\beta_3||\} \leq K\}$. Then

$$\begin{aligned} &c_3^{\intercal}\sqrt{n}(\hat{\beta}_3 - \beta_3^*) = w_3(\hat{\Sigma}_3, \mathbb{G}_n, \beta_3^*), \quad c_3^{\intercal}\sqrt{n}(\hat{\beta}_3^{(b)} - \hat{\beta}_3) = w_3(\hat{\Sigma}_3^{(b)}, \mathbb{G}_n^{(b)}, \hat{\beta}_3), \\ &\text{and } c_3^{\intercal}\sqrt{n}(\hat{\beta}_3 - \beta_{3,n}^*) = w_3(\hat{\Sigma}_3, \sqrt{n}(\mathbb{P}_n - P_n), \beta_{3,n}^*). \end{aligned}$$

The limiting behavior of the above quantities can be obtained using the same arguments as those in Theorem 1.1.

Stage 2. For a given positive integer J, let $\Omega_J(\mathbb{R}^p)$ denote the set of functions $\mathbb{R}^p \to \{\{1\}, \ldots, \{J\}, \{1,2\}, \{1,3\}, \ldots, \{1,J\}, \ldots, \{1,\ldots,J\}\}$. For any $c_2 \in \mathbb{R}^{p_2}$, define the following functions.

1. $w_{2,1}: D_{p_2} \times l^\infty(\mathcal{F}_{2,1}) \times l^\infty(\mathcal{F}_{2,1}) \times \mathbb{R}^{p_3} \times \mathbb{R}^{p_2+p_3} \to \mathbb{R}$ is defined as

$$w_{2,1}(\Sigma_2, \omega, \mu, \nu_3, \beta) \triangleq \omega\left[c_2^\intercal \Sigma_2^{-1} B_2\big(Y_2 + \max_{i=1,\ldots,K_3} H_{3,1}^\intercal \beta_{3,i} - B_2^\intercal \beta_2\big)\right] + \mu\left[c_2^\intercal \Sigma_2^{-1} B_2\Big(\max_{i\in\mathcal{A}_3^*(H_{3,1})} H_{3,1}^\intercal \nu_{3,i}\Big) 1_{\#\mathcal{A}_3^*(H_{3,1})=1}\right],$$

where $\mathcal{F}_{2,1} = \Big\{f(b_2, y_3, h_{3,1}) = a^\intercal b_2\big(y_2 + \max_{i=1,\ldots,K_3} h_{3,1}^\intercal \beta_{3,i} - b_2^\intercal \beta_2\big) + a'^\intercal b_2 1_{\#\mathcal{A}_3^*(h_{3,1})=1} \big(\max_{i\in\mathcal{A}_3^*(h_{3,1})} h_{3,1}^\intercal \nu_{3,i}\big) : a, a' \in \mathbb{R}^{p_2}, \beta = (\beta_2^\intercal, \beta_3^\intercal)^\intercal \in \mathbb{R}^{p_2+p_3}, \nu_3 = (\nu_{3,1}^\intercal, \ldots, \nu_{3,K_3}^\intercal)^\intercal \in \mathbb{R}^{p_3}, \max\{||a||, ||a'||, ||\beta||, ||\nu_3||\} \le K\Big\}$.

2. $w_{2,2}: D_{p_2} \times l^\infty(\mathcal{F}_{2,2}) \times \mathbb{R}^{p_3} \times \mathbb{R}^{p_3} \to \mathbb{R}$ is defined as

$$w_{2,2}(\Sigma_2, \mu, \nu_3, \gamma_3) \triangleq \mu\left[c_2^\intercal \Sigma_2^{-1} B_2\left(\max_{i\in\mathcal{A}_3^*(H_{3,1})} \big(H_{3,1}^\intercal \nu_{3,i} + H_{3,1}^\intercal \gamma_{3,i}\big) - \max_{i\in\mathcal{A}_3^*(H_{3,1})} H_{3,1}^\intercal \gamma_{3,i}\right) 1_{\#\mathcal{A}_3^*(H_{3,1})>1}\right],$$

where $\mathcal{F}_{2,2} = \Big\{f(b_2, h_{3,1}) = a^\intercal b_2\big(\max_{i\in\mathcal{A}_3^*(h_{3,1})}(h_{3,1}^\intercal \nu_{3,i} + h_{3,1}^\intercal \gamma_{3,i}) - \max_{i\in\mathcal{A}_3^*(h_{3,1})} h_{3,1}^\intercal \gamma_{3,i}\big) \times 1_{\#\mathcal{A}_3^*(h_{3,1})>1} : a \in \mathbb{R}^{p_2}, \gamma_3, \nu_3 \in \mathbb{R}^{p_3}, \max\{||a||, ||\nu_3||\} \le K\Big\}$.

3. $\rho_{2,0}: D_{p_2} \times l^\infty(\widetilde{\mathcal{F}}_{2,0}) \times \Omega_{K_3}(\mathbb{R}^{p_3/K_3}) \times \mathbb{R}^{p_3} \times \mathbb{R}^{p_3} \to \mathbb{R}$ is defined as

$$\rho_{2,0}(\Sigma_2, \mu, \mathcal{A}_3, \nu_3, \gamma_3) \triangleq \mu\left[c_2^\intercal \Sigma_2^{-1} B_2\left(\max_{i\in\mathcal{A}_3(H_{3,1})} \big(H_{3,1}^\intercal \nu_{3,i} + H_{3,1}^\intercal \gamma_{3,i}\big) - \max_{i\in\mathcal{A}_3(H_{3,1})} H_{3,1}^\intercal \gamma_{3,i} - \max_{i\in\mathcal{A}_3^*(H_{3,1})} \big(H_{3,1}^\intercal \nu_{3,i} + H_{3,1}^\intercal \gamma_{3,i}\big) + \max_{i\in\mathcal{A}_3^*(H_{3,1})} H_{3,1}^\intercal \gamma_{3,i}\right) 1_{\#\mathcal{A}_3^*(H_{3,1})>1}\right],$$

where $\widetilde{\mathcal{F}}_{2,0} = \Big\{f(b_2, h_{3,1}) = a^\intercal b_2\big(\max_{i\in\mathcal{A}_3(h_{3,1})}(h_{3,1}^\intercal \nu_{3,i} + h_{3,1}^\intercal \gamma_{3,i}) - \max_{i\in\mathcal{A}_3(h_{3,1})} h_{3,1}^\intercal \gamma_{3,i} - \max_{i\in\mathcal{A}_3^*(h_{3,1})}(h_{3,1}^\intercal \nu_{3,i} + h_{3,1}^\intercal \gamma_{3,i}) + \max_{i\in\mathcal{A}_3^*(h_{3,1})} h_{3,1}^\intercal \gamma_{3,i}\big) 1_{\#\mathcal{A}_3^*(h_{3,1})>1} : a \in \mathbb{R}^{p_2}, \gamma_3, \nu_3 \in \mathbb{R}^{p_3}, \max\{||a||, ||\nu_3||\} \le K, \mathcal{A}_3 \in \Omega_{K_3}(\mathbb{R}^{p_3/K_3})\Big\}$ and note that p_3/K_3 is the dimension of $\beta_{3,i}^*$ for $i = 1, \ldots, K_3$.

4. $\rho_{2,1} : D_{p_2} \times D^k_{p_3/K_3} \times \ldots \times D^k_{p_3/K_3} \times l^\infty(\widetilde{\mathcal{F}}_{2,1}) \times \Omega_{K_3}(\mathbb{R}^{p_3/K_3}) \times \mathbb{R}^{p_3} \times \mathbb{R}^{p_3} \times \mathbb{R}^{p_3} \times \mathbb{R} \to \mathbb{R}$ is defined as

$$\rho_{2,1}(\Sigma_2, \zeta_{3,1}, \ldots, \zeta_{3,K_3}, \mu, \mathcal{A}_3, \nu_3, \eta_3, \gamma_3, \lambda) \triangleq \mu\Bigg[c_2^\intercal \Sigma_2^{-1} B_2\Big(\max_{i\in\mathcal{A}_3(H_{3,1})} \left(H_{3,1}^\intercal \nu_{3,i} + H_{3,1}^\intercal \gamma_{3,i}\right)$$
$$- \max_{i\in\mathcal{A}_3(H_{3,1})} H_{3,1}^\intercal \gamma_{3,i}\Big) \times \left(1_{\min_i \frac{[H_{3,1}^\intercal \nu_{3,i} + H_{3,1}^\intercal \eta_{3,i} - \max_{j\neq i}(H_{3,1}^\intercal \nu_{3,j} + H_{3,1}^\intercal \eta_{3,j})]^2}{H_{3,1}^\intercal \zeta_{3,i} H_{3,1}} \leq \lambda} - 1_{\#\mathcal{A}_3^*(h_{3,1})>1}\right)\Bigg],$$

where $\widetilde{\mathcal{F}}_{2,1} = \Big\{f(b_1, h_{3,1}) = a^\intercal b_2 \left(\max_{i\in\mathcal{A}_3(h_{3,1})} \left(h_{3,1}^\intercal \nu_{3,i} + h_{3,1}^\intercal \gamma_{3,i}\right) - \max_{i\in\mathcal{A}_3(h_{3,1})} h_{3,1}^\intercal \gamma_{3,i}\right) \times \Big(1_{\min_i \frac{[h_{3,1}^\intercal \nu_{3,i} + h_{3,1}^\intercal \eta_{3,i} - \max_{j\neq i}(h_{3,1}^\intercal \nu_{3,j} + h_{3,1}^\intercal \eta_{3,j})]^2}{h_{3,1}^\intercal \zeta_{3,i} h_{3,1}} \leq \lambda} - 1_{\#\mathcal{A}_3^*(h_{3,1})>1}\Big) : a \in \mathbb{R}^{p_2}, \nu_3, \eta_3, \gamma_3 \in \mathbb{R}^{p_3}, \max\{||a||, ||\nu_3||\} \leq K, \mathcal{A}_3 \in \Omega_{K_3}(\mathbb{R}^{p_3/K_3}), \lambda \in \mathbb{R}, \zeta_{3,i} \in D^k_{p_3/K_3} \text{ for } i = 1, \ldots, K_3\Big\}$.

5. $\rho_{2,2} : D_{p_2} \times l^\infty(\widetilde{\mathcal{F}}_{2,2}) \times \mathbb{R}^{p_3} \times \mathbb{R}^{p_3} \to \mathbb{R}$, defined as

$$\rho_{2,2}(\Sigma_2, \mu, \nu_3, \eta_3) \triangleq \mu\Bigg[c_2^\intercal \Sigma_2^{-1} B_2\Big(\max_{i=1,\ldots,K_3} \left(H_{3,1}^\intercal \nu_{3,i} + H_{3,1}^\intercal \eta_{3,i}\right)$$
$$- \max_{i=1,\ldots,K_3} H_{3,1}^\intercal \eta_{3,i} - \max_{i\in\mathcal{A}_3^*(H_{3,1})} H_{3,1}^\intercal \nu_{3,i}\Big) 1_{\#\mathcal{A}_3^*(H_{3,1})=1}\Bigg],$$

where $\widetilde{\mathcal{F}}_{2,2} = \Big\{f(b_2, h_{3,1}) = a^\intercal b_2\big(\max_{i=1,\ldots,K_3} \left(h_{3,1}^\intercal \nu_{3,i} + h_{3,1}^\intercal \eta_{3,i}\right) - \max_{i=1,\ldots,K_3} h_{3,1}^\intercal \eta_{3,i} - \max_{i\in\mathcal{A}_3^*(h_{3,1})} h_{3,1}^\intercal \nu_{3,i}\big) 1_{\#\mathcal{A}_3^*(h_{3,1})=1} : \nu_3, \eta_3 \in \mathbb{R}^{p_3}, \max\{||a||, ||\nu_3||\} \leq K\Big\}$.

Define $v_{c_2} : D_{p_2} \times D^k_{p_3/K_3} \times \ldots \times D^k_{p_3/K_3} \times l^\infty(\mathcal{F}_{2,1}) \times l^\infty(\bar{\mathcal{F}}_2) \times \Omega_{K_3}(\mathbb{R}^{p_3/K_3}) \times \mathbb{R}^{p_2+p_3} \times \mathbb{R}^{p_3} \times \mathbb{R}^{p_3} \times \mathbb{R}^{p_3} \times \mathbb{R} \to \mathbb{R}$ as

$$\begin{aligned}
&v_{c_2}(\Sigma_2, \zeta_{3,1}, \ldots, \zeta_{3,K_3}, \omega, \mu, \mathcal{A}_3, \beta, \nu_3, \eta_3, \gamma_3) \\
&\quad = w_{2,1}(\Sigma_2, \omega, \mu, \nu_3, \beta) + \rho_{2,2}(\Sigma_2, \mu, \nu_3, \eta_3) - \rho_{2,1}(\Sigma_2, \zeta_{3,1}, \ldots, \zeta_{3,K_3}, \mu, \mathcal{A}_3, \nu_3, \eta_3, \eta_3, \lambda_n) \\
&\qquad + w_{2,2}(\Sigma_2, \mu, \nu_3, \gamma_3) + \rho_{2,0}(\Sigma_2, \mu, \mathcal{A}_3, \nu_3, \gamma_3) \\
&\qquad + \rho_{2,1}(\Sigma_2, \zeta_{3,1}, \ldots, \zeta_{3,K_3}, \mu, \mathcal{A}_3, \nu_3, \eta_3, \gamma_3, \lambda_n), \quad (19)
\end{aligned}$$

where $\bar{\mathcal{F}}_2 = \mathcal{F}_{2,1} \oplus \mathcal{F}_{2,2} \oplus \widetilde{\mathcal{F}}_{1,0} \oplus \widetilde{\mathcal{F}}_{1,1} \oplus \widetilde{\mathcal{F}}_{1,2}$, and $\oplus$ denotes element-wise addition. Then

$$
\begin{aligned}
c_2^\intercal \sqrt{n}(\hat{\beta}_2 - \beta_2^*) =& v_{c_2}(\hat{\Sigma}_2, \hat{\zeta}_{3,1}, \ldots, \hat{\zeta}_{3,K_3}, \mathbb{G}_n, \mathbb{P}_n, \widetilde{\mathcal{A}}_3, (\beta_2^{*\intercal}, \beta_3^{*\intercal})^\intercal, \sqrt{n}(\hat{\beta}_3 - \beta_3^*), \sqrt{n}\beta_3^*, \sqrt{n}\beta_3^*), \\
c_2^\intercal \sqrt{n}(\hat{\beta}_2 - \beta_{2,n}^*) =& v_{c_2}(\hat{\Sigma}_2, \hat{\zeta}_{3,1}, \ldots, \hat{\zeta}_{3,K_3}, \\
& \sqrt{n}(\mathbb{P}_n - P_n), \mathbb{P}_n, \widetilde{\mathcal{A}}_{3,n}, (\beta_{2,n}^{*\intercal}, \beta_{3,n}^{*\intercal})^\intercal, \sqrt{n}(\hat{\beta}_3 - \beta_{3,n}^*), \sqrt{n}\beta_{3,n}^*, \sqrt{n}\beta_{3,n}^*).
\end{aligned}
$$

The upper bound $\mathcal{U}_2(c_2)$ under P is

$$
\mathcal{U}_2(c_2) = \sup_{\gamma_3 \in \mathbb{R}^{p_3}} v_{c_2}\left(\hat{\Sigma}_2, \hat{\zeta}_{3,1}, \ldots, \hat{\zeta}_{3,K_3}, \mathbb{G}_n, \mathbb{P}_n, \widetilde{\mathcal{A}}_3, (\beta_2^{*\intercal}, \beta_3^{*\intercal})^\intercal, \sqrt{n}(\hat{\beta}_3 - \beta_3^*), \sqrt{n}\beta_3^*, \gamma_3\right).
$$

The upper bound $\mathcal{U}_2(c_2)$ under P_n is

$$
\mathcal{U}_2(c_2) = \sup_{\gamma_3 \in \mathbb{R}^{p_3}} v_{c_2}\left(\hat{\Sigma}_2, \hat{\zeta}_{3,1}, \ldots, \hat{\zeta}_{3,K_3}, \sqrt{n}(\mathbb{P}_n - P_n), \mathbb{P}_n, \widetilde{\mathcal{A}}_{3,n}, (\beta_{2,n}^{*\intercal}, \beta_{3,n}^{*\intercal})^\intercal, \sqrt{n}(\hat{\beta}_3 - \beta_{3,n}^*), \sqrt{n}\beta_{3,n}^*, \gamma_3\right).
$$

And the bootstrap analog of the upper bound is

$$
\hat{\mathcal{U}}_2^{(b)}(c_2) = \sup_{\gamma_3 \in \mathbb{R}^{p_3}} v_{c_2}\left(\hat{\Sigma}_2^{(b)}, \hat{\zeta}_{3,1}^{(b)}, \ldots, \hat{\zeta}_{3,K_3}^{(b)}, \sqrt{n}(\hat{\mathbb{P}}_n^{(b)} - \mathbb{P}_n), \hat{\mathbb{P}}_n^{(b)}, \widetilde{\mathcal{A}}_3^{(b)}, (\hat{\beta}_2^\intercal, \hat{\beta}_3^\intercal)^\intercal, \sqrt{n}(\hat{\beta}_3^{(b)} - \hat{\beta}_3), \sqrt{n}\hat{\beta}_3, \gamma_3\right).
$$

Results for $t = 2$ in Theorems 2.1 and 2.2 rely on the desired continuity of $w_{2,1}$ and $w_{2,2}$ and negligibility of the error terms $\rho_{2,0}$, $\rho_{2,1}$ and $\rho_{2,2}$, which can be proved using similar arguments as those in Section 1.2 and are omitted. Below we show the negligibility of $\rho_{2,0}$ as an example.

Theorem 2.4. *Assume (A1)-(A3). Then*

1. $\sup_{\gamma_3 \in \mathbb{R}^{p_3}} |\rho_{2,0}(\hat{\Sigma}_2, \mathbb{P}_n, \widetilde{\mathcal{A}}_3, \sqrt{n}(\hat{\beta}_3 - \beta_3^*), \gamma_3)| \to_P 0$*; and*

2. $\sup_{\gamma_3 \in \mathbb{R}^{p_3}} |\rho_{2,0}(\hat{\Sigma}_2^{(b)}, \hat{\mathbb{P}}_n^{(b)}, \widetilde{\mathcal{A}}_3^{(b)}, \sqrt{n}(\hat{\beta}_3^{(b)} - \hat{\beta}_3), \gamma_3)| \to_{P_M} 0$ *a.s. P.*

If, in addition (A4) holds, then

3. $\sup_{\gamma_3\in\mathbb{R}^{p_3}} |\rho_{2,0}(\hat{\Sigma}_2, \mathbb{P}_n, \widetilde{\mathcal{A}}_{3,n}, \sqrt{n}(\hat{\beta}_3 - \beta^*_{3,n}), \gamma_3)| \to_{P_n} 0$.

Proof. For any probability measure $\mu \in l^\infty(\widetilde{\mathcal{F}}_{2,0})$ and positive definite Σ_2, we have

$$
\begin{aligned}
|\rho_{2,0}(\Sigma_2, \mu, \mathcal{A}_3, \nu_3, \gamma_3)| &= \Big|\mu\Big[c_2^\intercal \Sigma_2^{-1} B_2 \Big(\max_{i\in\mathcal{A}_3(H_{3,1})} \big(H_{3,1}^\intercal \nu_{3,i} + H_{3,1}^\intercal \gamma_{3,i}\big) - \max_{i\in\mathcal{A}_3(H_{3,1})} H_{3,1}^\intercal \gamma_{3,i} \\
&\quad - \max_{i\in\mathcal{A}_3^*(H_{3,1})} \big(H_{3,1}^\intercal \nu_{3,i} + H_{3,1}^\intercal \gamma_{3,i}\big) + \max_{i\in\mathcal{A}_3^*(H_{3,1})} H_{3,1}^\intercal \gamma_{3,i}\Big) 1_{\#\mathcal{A}_3^*(H_{3,1})>1} 1_{\mathcal{A}_3(H_{3,1})\neq\mathcal{A}_3^*(H_{3,1})}\Big]\Big| \\
&\leq \mu\left[|c_2^\intercal \Sigma_2^{-1} B_2| \left(\max_{i\in\mathcal{A}_3(H_{3,1})} |H_{3,1}^\intercal \nu_{3,i}| + \max_{i\in\mathcal{A}_3^*(H_{3,1})} |H_{3,1}^\intercal \nu_{3,i}|\right) 1_{\mathcal{A}_3(H_{3,1})\neq\mathcal{A}_3^*(H_{3,1})}\right] \\
&\leq K||\nu_3||\, \mu\left[||B_2||\, ||H_{3,1}|| 1_{\mathcal{A}_3(H_{3,1})\neq\mathcal{A}_3^*(H_{3,1})}\right]
\end{aligned}
$$

for a sufficiently large constant K, where the first inequality follows from the fact that $|\max_i(a_i+b_i) - \max_i b_i| \leq \max_i |a_i|$ for any vectors a and b of the same dimension, and the second inequality follows from Cauchy-Schwarz inequality and the fact that $\max_{i=1,\ldots,K_3} ||\nu_{3,i}|| \leq ||\nu_3||$.

To prove result 1, note that $\hat{\Sigma}_2$ is positive definite for sufficiently large n. Thus

$$
|\rho_{2,0}(\hat{\Sigma}_2, \mathbb{P}_n, \widetilde{\mathcal{A}}_3, \sqrt{n}(\hat{\beta}_3 - \beta_3^*), \gamma_3)| \leq K||\sqrt{n}(\hat{\beta}_3 - \beta_3^*)||\mathbb{P}_n\left[||B_2||\, ||H_{3,1}|| 1_{\widetilde{\mathcal{A}}_3(H_{3,1})\neq\mathcal{A}_3^*(H_{3,1})}\right].
$$

Below we show that $\mathbb{P}_n[||B_2||\, ||H_{3,1}|| 1_{\widetilde{\mathcal{A}}_3(H_{3,1})\neq\mathcal{A}_3^*(H_{3,1})}] = o_P(1)$. Define $\Delta_3(h_{3,1}) \triangleq \#\{\hat{\mathcal{A}}_3(h_{3,1})\Delta\mathcal{A}_3^*(h_{3,1})\}$. Then $1_{\widetilde{\mathcal{A}}_3(H_{3,1})\neq\mathcal{A}_3^*(H_{3,1})} \leq 1_{\Delta_3(H_{3,1})>0}$. For any $\delta > 0$, choose ϵ sufficiently small so that $P||B_2||\, ||H_{3,1}|| 1_{H_{3,1}\notin B_{3,\epsilon}} \leq \eta/2$, where $B_{3,\epsilon}$ is as defined in Lemma 2.7. Then

$$
\begin{aligned}
\mathbb{P}_n||B_2||\, ||H_{3,1}|| 1_{\widetilde{\mathcal{A}}_3(H_{3,1})\neq\mathcal{A}_3^*(H_{3,1})} &\leq \mathbb{P}_n||B_2||\, ||H_{3,1}|| 1_{\Delta_3(H_{3,1})>0} \\
&\leq \mathbb{P}_n||B_2||\, ||H_{3,1}|| \sup_{h_{3,1}\in B_{3,\epsilon}} 1_{\Delta_3(h_{3,1})>0} + \mathbb{P}_n||B_2||\, ||H_{3,1}|| 1_{h_{3,1}\notin B_{3,\epsilon}} < \delta
\end{aligned}
$$

with probability tending to one by appeal to Lemma 2.7, the LLN, and Slutsky's theorem. This together with the fact that $||\sqrt{n}(\hat{\beta}_3 - \beta_3^*)|| = O_P(1)$ completes the proof of result 1.

Similar arguments can be used to prove results 2 and 3, and are omitted. $\square$

Next we will provide proof sketches for Theorems 2.1 and 2.2 for $t = 1$.

Stage 1. Let $c(j)$ be the j-th column of $I_{p_2\times p_2}$. Note that by definition

$$\begin{aligned}
\widetilde{\mathbb{V}}_{2,n}(\gamma_3) &= \left(v_{c(j)}(\hat{\Sigma}_2, \hat{\zeta}_{3,1}, \dots, \hat{\zeta}_{3,K_3}, \mathbb{G}_n, \mathbb{P}_n, \tilde{\mathcal{A}}_3, (\beta_2^{*\intercal}, \beta_3^{*\intercal})^\intercal, \sqrt{n}(\hat{\beta}_3 - \beta_3^*), \sqrt{n}\beta_3^*, \gamma_3)\right)_{j=1}^{p_2} \\
&= \begin{pmatrix} v_{c(1)}(\hat{\Sigma}_2, \hat{\zeta}_{3,1}, \dots, \hat{\zeta}_{3,K_3}, \mathbb{G}_n, \mathbb{P}_n, \tilde{\mathcal{A}}_3, (\beta_2^{*\intercal}, \beta_3^{*\intercal})^\intercal, \sqrt{n}(\hat{\beta}_3 - \beta_3^*), \sqrt{n}\beta_3^*, \gamma_3) \\ \cdots \\ v_{c(p_2)}(\hat{\Sigma}_2, \hat{\zeta}_{3,1}, \dots, \hat{\zeta}_{3,K_3}, \mathbb{G}_n, \mathbb{P}_n, \tilde{\mathcal{A}}_3, (\beta_2^{*\intercal}, \beta_3^{*\intercal})^\intercal, \sqrt{n}(\hat{\beta}_3 - \beta_3^*), \sqrt{n}\beta_3^*, \gamma_3) \end{pmatrix}.
\end{aligned}$$

The analog of $\widetilde{\mathbb{V}}_{2,n}$ under local alternatives P_n is

$$\widetilde{\mathbb{V}}_{2,n}^{P_n}(\gamma_3) = \left(v_{c(j)}(\hat{\Sigma}_2, \hat{\zeta}_{3,1}, \dots, \hat{\zeta}_{3,K_3}, \sqrt{n}(\mathbb{P}_n - P_n), \mathbb{P}_n, \tilde{\mathcal{A}}_{3,n}, (\beta_{2,n}^{*\intercal}, \beta_{3,n}^{*\intercal})^\intercal, \sqrt{n}(\hat{\beta}_3 - \beta_{3,n}^*), \sqrt{n}\beta_{3,n}^*, \gamma_3)\right)_{j=1}^{p_2}. \tag{20}$$

And the bootstrap analog of $\widetilde{\mathbb{V}}_{2,n}$ is

$$\widetilde{\mathbb{V}}_{2,n}^{(b)}(\gamma_3) = \left(v_{c(j)}(\hat{\Sigma}_2^{(b)}, \hat{\zeta}_{3,1}^{(b)}, \dots, \hat{\zeta}_{3,K_3}^{(b)}, \sqrt{n}(\mathbb{P}_n^{(b)} - \mathbb{P}_n), \mathbb{P}_n^{(b)}, \tilde{\mathcal{A}}_3^{(b)}, (\hat{\beta}_2^\intercal, \hat{\beta}_3^\intercal)^\intercal, \sqrt{n}(\hat{\beta}_3^{(b)} - \hat{\beta}_3), \sqrt{n}\hat{\beta}_3, \gamma_3)\right)_{j=1}^{p_2}.$$

Let $\Pi_{p_2} l^\infty(\mathbb{R}^{p_3})$ denote the set of bounded p_2 vector-valued functions on $\mathbb{R}^{p_3}$ equipped with the sup norm (over $\{1, \dots, p_2\} \times \mathbb{R}^{p_3}$). For any $c_1 \in \mathbb{R}^{p_1}$, we define the following functions.

1. $w_{1,1} : D_{p_1} \times l^\infty(\mathcal{F}_{1,1}) \times \mathbb{R}^{p_1+p_2} \to \mathbb{R}$ is defined as

$$w_{1,1}(\Sigma_1, \omega, \beta) = \omega\left[c_1^\intercal \Sigma_1^{-1} B_1 \left(Y_1 + \max_{i=1,\dots,K_2} H_{2,1}^\intercal \beta_{2,i} - B_1^\intercal \beta_1\right)\right], \tag{21}$$

where $\mathcal{F}_{1,1} = \{f(b_1, y_1, h_{2,1}) = a^\intercal b_1 \left(y_1 + \max_{i=1,\dots,K_2} h_{2,1}^\intercal \beta_{2,i} - b_1^\intercal \beta_1\right) : a \in \mathbb{R}^{p_1}, \beta = (\beta_1^\intercal, \beta_2^\intercal)^\intercal \in \mathbb{R}^{p_1+p_2}, \max\{||a||, ||\beta||\} \leq K\}$.

2. $w_{1,2} : D_{p_1} \times l^\infty(\mathcal{F}_{1,2}) \times \Pi_{p_2} l^\infty(\mathbb{R}^{p_3}) \times \mathbb{R}^{p_3} \to \mathbb{R}$ is defined as

$$w_{1,2}(\Sigma_1, \mu, \phi, \gamma_3) = \mu \left[c_1^\intercal \Sigma_1^{-1} B_1 \max_{i \in \mathcal{A}_2^*(H_{2,1})} H_{2,1}^\intercal \phi_i(\gamma_3) 1_{\#\mathcal{A}_2^*(H_{2,1})=1} \right] \tag{22}$$

where $\mathcal{F}_{1,2} = \{f(b_1, h_{2,1}) = a^\intercal b_1 \max_{i \in \mathcal{A}_2^*(h_{2,1})} h_{2,1}^\intercal \phi_i(\gamma_3) 1_{\#\mathcal{A}_2^*(h_{2,1})=1} : a \in \mathbb{R}^{p_1}, \phi = (\phi_1^\intercal, \ldots, \phi_{K_2}^\intercal)^\intercal \in \Pi_{p_2} l^\infty(\mathbb{R}^{p_3}), \gamma_3 \in \mathbb{R}^{p_3}, ||a|| \leq K\}$.

3. $w_{1,3} : D_{p_1} \times l^\infty(\mathcal{F}_{1,3}) \times \Pi_{p_2} l^\infty(\mathbb{R}^{p_3}) \times \mathbb{R}^{p_2} \times \mathbb{R}^{p_3} \to \mathbb{R}$ is defined as

$$\begin{aligned} w_{1,3}(\Sigma_1, \mu, \phi, \gamma_2, \gamma_3) = \mu \Bigg[& c_1^\intercal \Sigma_1^{-1} B_1 \\ & \times \left(\max_{i \in \mathcal{A}_2^*(H_{2,1})} \left(H_{2,1}^\intercal \phi_i(\gamma_3) + H_{2,1}^\intercal \gamma_{2,i} \right) - \max_{i \in \mathcal{A}_2^*(H_{2,1})} H_{2,1}^\intercal \gamma_{2,i} \right) 1_{\#\mathcal{A}_2^*(H_{2,1})>1} \Bigg] \end{aligned} \tag{23}$$

where $\mathcal{F}_{1,3} = \{f(b_1, h_{2,1}) = a^\intercal b_1 \left(\max_{i \in \mathcal{A}_2^*(h_{2,1})} \left(h_{2,1}^\intercal \phi(\gamma_3) + h_{2,1}^\intercal \gamma_{2,i} \right) - \max_{i \in \mathcal{A}_2^*(h_{2,1})} h_{2,1}^\intercal \gamma_{2,i} \right)$ $1_{\#\mathcal{A}_2^*(h_{2,1})>1} : a \in \mathbb{R}^{p_1}, \phi = (\phi_1^\intercal, \ldots, \phi_{K_2}^\intercal)^\intercal \in \Pi_{p_2} l^\infty(\mathbb{R}^{p_3}), \gamma_2 \in \mathbb{R}^{p_2}, \gamma_3 \in \mathbb{R}^{p_3}, ||a|| \leq K\}$.

4. $\rho_{1,0} : D_{p_1} \times l^\infty(\widetilde{\mathcal{F}}_{1,0}) \times \Pi_{p_2} l^\infty(\mathbb{R}^{p_3}) \times \Omega_{K_2}(\mathbb{R}^{p_2/K_2}) \times \mathbb{R}^{p_3} \to \mathbb{R}$ is defined as

$$\rho_{1,0}(\Sigma_1, \mu, \phi, \mathcal{A}_2, \gamma_3) = \mu \left[c_1^\intercal \Sigma_1^{-1} B_1 \left(\max_{i \in \mathcal{A}_2(H_{2,1})} H_{2,1}^\intercal \phi_i(\gamma_3) - \max_{i \in \mathcal{A}_2^*(H_{2,1})} H_{2,1}^\intercal \phi_i(\gamma_3) \right) 1_{\#\mathcal{A}_2^*(H_{2,1})=1} \right], \tag{24}$$

where $\widetilde{\mathcal{F}}_{1,0} = \{f(b_1, h_{2,1}) = a^\intercal b_1 \left(\max_{i \in \mathcal{A}_2(h_{2,1})} h_{2,1}^\intercal \phi_i(\gamma_3) - \max_{i \in \mathcal{A}_2^*(h_{2,1})} h_{2,1}^\intercal \phi_i(\gamma_3) \right) 1_{\#\mathcal{A}_2^*(h_{2,1})=1} :$ $a \in \mathbb{R}^{p_1}, ||a|| \leq K, \phi = (\phi_1^\intercal, \ldots, \phi_{K_2}^\intercal)^\intercal \in \Pi_{p_2} l^\infty(\mathbb{R}^{p_3}), \mathcal{A}_2 \in \Omega_{K_2}(\mathbb{R}^{p_2/K_2}), \gamma_3 \in \mathbb{R}^{p_3}\}$.

5. $\rho_{1,1} : D_{p_1} \times l^\infty(\widetilde{\mathcal{F}}_{1,1}) \times \Pi_{p_2} l^\infty(\mathbb{R}^{p_3}) \times \Omega_{K_2}(\mathbb{R}^{p_2/K_2}) \times \mathbb{R}^{p_2} \times \mathbb{R}^{p_3} \to \mathbb{R}$ is defined as

$$\begin{aligned} \rho_{1,1}(\Sigma_1, \mu, \phi, \mathcal{A}_2, \gamma_2, \gamma_3) = \mu \Bigg[& c_1^\intercal \Sigma_1^{-1} B_1 \Bigg(\max_{i \in \mathcal{A}_2(H_{2,1})} \left(H_{2,1}^\intercal \phi_i(\gamma_3) + H_{2,1}^\intercal \gamma_{2,i} \right) - \max_{i \in \mathcal{A}_2(H_{2,1})} H_{2,1}^\intercal \gamma_{2,i} \\ & - \max_{i \in \mathcal{A}_2^*(H_{2,1})} \left(H_{2,1}^\intercal \phi_i(\gamma_3) + H_{2,1}^\intercal \gamma_{2,i} \right) + \max_{i \in \mathcal{A}_2^*(H_{2,1})} H_{2,1}^\intercal \gamma_{2,i} \Bigg) 1_{\#\mathcal{A}_2^*(H_{2,1})>1} \Bigg], \end{aligned} \tag{25}$$

where $\widetilde{\mathcal{F}}_{1,1} = \{f(b_1, h_{2,1}) = a^\intercal b_1 \Big(\max_{i \in \mathcal{A}_2(h_{2,1})} \left(h_{2,1}^\intercal \phi_i(\gamma_3) + h_{2,1}^\intercal \gamma_{2,i} \right) - \max_{i \in \mathcal{A}_2(h_{2,1})} h_{2,1}^\intercal \gamma_{2,i}$

$-\max_{i\in\mathcal{A}_2^*(h_{2,1})}\left(h_{2,1}^\intercal\phi_i(\gamma_3)+h_{2,1}^\intercal\gamma_{2,i}\right)+\max_{i\in\mathcal{A}_2^*(h_{2,1})}h_{2,1}^\intercal\gamma_{2,i}\Big)1_{\#\mathcal{A}_2^*(h_{2,1})>1} : a\in\mathbb{R}^{p_1}, ||a||\leq K, \phi=(\phi_1^\intercal,\ldots,\phi_{K_2}^\intercal)^\intercal\in\Pi_{p_2}l^\infty(\mathbb{R}^{p_3}), \mathcal{A}_2\in\Omega_{K_2}(\mathbb{R}^{p_2/K_2}), \gamma_2\in\mathbb{R}^{p_2}, \gamma_3\in\mathbb{R}^{p_3}\}$.

6. $\rho_{1,2}: D_{p_1}\times D^k_{p_2/K_2}\times\ldots\times D^k_{p_2/K_2}\times l^\infty(\widetilde{\mathcal{F}}_{1,2})\times\Pi_{p_2}l^\infty(\mathbb{R}^{p_3})\times\Omega_{K_2}(\mathbb{R}^{p_2/K_2})\times\Omega_{K_2}(\mathbb{R}^{p_2/K_2})\times\mathbb{R}^{p_2}\times\mathbb{R}^{p_2}\times\mathbb{R}^{p_2}\times\mathbb{R}^{p_3}\times\mathbb{R}\to\mathbb{R}$ is defined as

$$\begin{aligned}&\rho_{1,2}(\Sigma_1,\zeta_{2,1},\ldots,\zeta_{2,K_2},\mu,\phi,\mathcal{A}_2,\mathcal{A}_2',\nu_2,\eta_2,\gamma_2,\gamma_3,\lambda)\\ &=\mu\Bigg[c_1^\intercal\Sigma_1^{-1}B_1\Big(\max_{i\in\mathcal{A}_2'(H_{2,1})}\left(H_{2,1}^\intercal\phi_i(\gamma_3)+H_{2,1}^\intercal\gamma_{2,i}\right)-\max_{i\in\mathcal{A}_2'(H_{2,1})}H_{2,1}^\intercal\gamma_{2,i}-\max_{i\in\mathcal{A}_2(H_{2,1})}H_{2,1}^\intercal\phi_i(\gamma_3)\Big)\\ &\qquad\times\left(1_{\min_i\frac{(H_{2,1}^\intercal\nu_{2,i}+H_{2,1}^\intercal\eta_{2,i}-\max_{j\neq i}(H_{2,1}^\intercal\nu_{2,j}+H_{2,1}^\intercal\eta_{2,j}))^2}{H_{2,1}^\intercal\zeta_{2,i}H_{2,1}}\leq\lambda}-1_{\#\mathcal{A}_2^*(H_{2,1})>1}\right)\Bigg],\end{aligned}\tag{26}$$

where $\widetilde{\mathcal{F}}_{1,2}=\{f(b_1,h_{2,1})=a^\intercal b_1\Big(\max_{i\in\mathcal{A}_2'(h_{2,1})}\left(h_{2,1}^\intercal\phi_i(\gamma_3)+h_{2,1}^\intercal\gamma_{2,i}\right)-\max_{i\in\mathcal{A}_2'(h_{2,1})}h_{2,1}^\intercal\gamma_{2,i}-\max_{i\in\mathcal{A}_2(h_{2,1})}h_{2,1}^\intercal\phi_i(\gamma_3)\Big)\times\left(1_{\min_i\frac{(h_{2,1}^\intercal\nu_{2,i}+h_{2,1}^\intercal\eta_{2,i}-\max_{j\neq i}(h_{2,1}^\intercal\nu_{2,j}+h_{2,1}^\intercal\eta_{2,j}))^2}{h_{2,1}^\intercal\zeta_{2,i}h_{2,1}}\leq\lambda}-1_{\#\mathcal{A}_2^*(h_{2,1})>1}\right) : a\in\mathbb{R}^{p_1}, \phi=(\phi_1^\intercal,\ldots,\phi_{K_2}^\intercal)^\intercal\in\Pi_{p_2}l^\infty(\mathbb{R}^{p_3}), \mathcal{A}_2,\mathcal{A}_2'\in\Omega_{K_2}(\mathbb{R}^{p_2/K_2}), \nu_2,\eta_2,\gamma_2\in\mathbb{R}^{p_2}, \gamma_3\in\mathbb{R}^{p_3}, \max\{||a||,||\nu_2||\}\leq K, \zeta_{2,i}\in D^k_{p_2/K_2}, i=1,\ldots,K_2\}$.

7. $\rho_{1,3}: D_{p_1}\times l^\infty(\widetilde{\mathcal{F}}_{1,3})\times\Pi_{p_2}l^\infty(\mathbb{R}^{p_3})\times\Omega_{K_2}(\mathbb{R}^{p_2/K_2})\times\mathbb{R}^{p_2}\times\mathbb{R}^{p_3}\to\mathbb{R}$ is defined as

$$\begin{aligned}\rho_{1,3}(\Sigma_1,\mu,\phi,\mathcal{A}_2,\eta_2,\eta_3)=\mu\Bigg[c_1^\intercal\Sigma_1^{-1}B_1\Big(&\max_{i=1,\ldots,K_2}\left(H_{2,1}^\intercal\phi_i(\eta_3)+H_{2,1}^\intercal\eta_{2,i}\right)\\ &-\max_{i=1,\ldots,K_2}H_{2,1}^\intercal\eta_{2,i}-\max_{i\in\mathcal{A}_2(H_{2,1})}H_{2,1}^\intercal\phi_i(\eta_3)\Big)1_{\#\mathcal{A}_2^*(H_{2,1})=1}\Bigg],\end{aligned}\tag{27}$$

where $\widetilde{\mathcal{F}}_{1,3}=\{f(b_1,h_{2,1})=a^\intercal b_1\Big(\max_{i=1,\ldots,K_2}\left(h_{2,1}^\intercal\phi_i(\eta_3)+h_{2,1}^\intercal\eta_{2,i}\right)-\max_{i=1,\ldots,K_2}h_{2,1}^\intercal\eta_{2,i}-\max_{i\in\mathcal{A}_2(h_{2,1})}h_{2,1}^\intercal\phi_i(\eta_3)\Big)1_{\#\mathcal{A}_2^*(h_{2,1})=1} : a\in\mathbb{R}^{p_1}, ||a||\leq K, \phi=(\phi_1^\intercal,\ldots,\phi_{K_2}^\intercal)^\intercal\in\Pi_{p_2}l^\infty(\mathbb{R}^{p_3}), \mathcal{A}_2\in\Omega_{K_2}(\mathbb{R}^{p_2/K_2}), \eta_2\in\mathbb{R}^{p_2}, \eta_3\in\mathbb{R}^{p_3}\}$.

Define $\tau: D_{p_1}\times D^k_{p_2/K_2}\times\ldots\times D^k_{p_2/K_2}\times l^\infty(\mathcal{F}_{1,1})\times l^\infty(\bar{\mathcal{F}}_1)\times\Pi_{p_2}l^\infty(\mathbb{R}^{p_3})\times\Omega_{K_2}(\mathbb{R}^{p_2/K_2})\times$

$\Omega_{K_2}(\mathbb{R}^{p_2/K_2}) \times \mathbb{R}^{p_1+p_2} \times \mathbb{R}^{p_2} \times \mathbb{R}^{p_2} \times \mathbb{R}^{p_2} \times \mathbb{R}^{p_3} \times \mathbb{R}^{p_3} \to \mathbb{R}$ as

$$\begin{aligned}
&\tau(\Sigma_1, \zeta_{2,1}, \dots, \zeta_{2,K_2}, \omega, \mu, \phi, \mathcal{A}_2, \mathcal{A}_2', \beta, \nu_2, \eta_2, \gamma_2, \eta_3, \gamma_3) \\
&\quad = w_{1,1}(\Sigma_1, \omega, \beta) + w_{1,2}(\Sigma_1, \mu, \phi, \gamma_3) + \rho_{1,0}(\Sigma_1, \mu, \phi, \mathcal{A}_2, \gamma_3) + \rho_{1,3}(\Sigma_1, \mu, \phi, \mathcal{A}_2, \eta_2, \eta_3) \\
&\qquad - \rho_{1,2}(\Sigma_1, \zeta_{2,1}, \dots, \zeta_{2,K_2}, \mu, \phi, \mathcal{A}_2, \mathcal{A}_2', \nu_2, \eta_2, \eta_2, \eta_3, \lambda_n) \\
&\qquad + w_{1,3}(\Sigma_1, \mu, \phi, \gamma_2, \gamma_3) + \rho_{1,1}(\Sigma_1, \mu, \phi, \mathcal{A}_2', \gamma_2, \gamma_3) \\
&\qquad + \rho_{1,2}(\Sigma_1, \zeta_{2,1}, \dots, \zeta_{2,K_2}, \mu, \phi, \mathcal{A}_2, \mathcal{A}_2', \nu_2, \eta_2, \gamma_2, \gamma_3, \lambda_n)
\end{aligned}$$

where $\bar{\mathcal{F}}_1 = \mathcal{F}_{1,2} \oplus \mathcal{F}_{1,3} \oplus \widetilde{\mathcal{F}}_{1,0} \oplus \widetilde{\mathcal{F}}_{1,1} \oplus \widetilde{\mathcal{F}}_{1,2} \oplus \widetilde{\mathcal{F}}_{1,3}$, and $\oplus$ denotes element-wise addition. Then

$$\begin{aligned}
c_1^\intercal \sqrt{n}(\hat{\beta}_1 - \beta_1^*) &= \tau\big(\hat{\Sigma}_1, \hat{\zeta}_{2,1}, \dots, \hat{\zeta}_{2,K_2}, \mathbb{G}_n, \mathbb{P}_n, \widetilde{\mathbb{V}}_{2,n}, \hat{\mathcal{A}}_2, \widetilde{\mathcal{A}}_2, \\
&\qquad (\beta_1^{*\intercal}, \beta_2^{*\intercal})^\intercal, \sqrt{n}(\hat{\beta}_2 - \beta_2^*), \sqrt{n}\beta_2^*, \sqrt{n}\beta_2^*, \sqrt{n}\beta_3^*, \sqrt{n}\beta_3^*\big). \\
c_1^\intercal \sqrt{n}(\hat{\beta}_1 - \beta_{1,n}^*) &= \tau\big(\hat{\Sigma}_1, \hat{\zeta}_{2,1}, \dots, \hat{\zeta}_{2,K_2}, \sqrt{n}(\mathbb{P}_n - P_n), \mathbb{P}_n, \widetilde{\mathbb{V}}_{2,n}^{P_n}, \hat{\mathcal{A}}_2, \widetilde{\mathcal{A}}_{2,n}, \\
&\qquad (\beta_{1,n}^{*\intercal}, \beta_{2,n}^{*\intercal})^\intercal, \sqrt{n}(\hat{\beta}_2 - \beta_{2,n}^*), \sqrt{n}\beta_{2,n}^*, \sqrt{n}\beta_{2,n}^*, \sqrt{n}\beta_{3,n}^*, \sqrt{n}\beta_{3,n}^*\big),
\end{aligned}$$

where $\widetilde{\mathbb{V}}_{2,n}^{P_n}$ is defined in (20). The upper bound $\mathcal{U}_1(c_1)$ under P is

$$\begin{aligned}
\mathcal{U}_1(c_1) = \sup_{\gamma_2 \in \mathbb{R}^{p_2}, \gamma_3 \in \mathbb{R}^{p_3}} \tau\Big(&\hat{\Sigma}_1, \hat{\zeta}_{2,1}, \dots, \hat{\zeta}_{2,K_2}, \mathbb{G}_n, \mathbb{P}_n, \widetilde{\mathbb{V}}_{2,n}, \hat{\mathcal{A}}_2, \widetilde{\mathcal{A}}_2, \\
&(\beta_1^{*\intercal}, \beta_2^{*\intercal})^\intercal, \sqrt{n}(\hat{\beta}_2 - \beta_2^*), \sqrt{n}\beta_2^*, \gamma_2, \sqrt{n}\beta_3^*, \gamma_3\Big).
\end{aligned}$$

The upper bound $\mathcal{U}_1(c_1)$ under P_n is

$$\begin{aligned}
\mathcal{U}_1(c_1) = \sup_{\gamma_2 \in \mathbb{R}^{p_2}, \gamma_3 \in \mathbb{R}^{p_3}} \tau\Big(&\hat{\Sigma}_1, \hat{\zeta}_{2,1}, \dots, \hat{\zeta}_{2,K_2}, \sqrt{n}(\mathbb{P}_n - P_n), \mathbb{P}_n, \widetilde{\mathbb{V}}_{2,n}^{P_n}, \hat{\mathcal{A}}_2, \widetilde{\mathcal{A}}_{2,n}, \\
&(\beta_{1,n}^{*\intercal}, \beta_{2,n}^{*\intercal})^\intercal, \sqrt{n}(\hat{\beta}_2 - \beta_{2,n}^*), \sqrt{n}\beta_{2,n}^*, \gamma_2, \sqrt{n}\beta_{3,n}^*, \gamma_3\Big).
\end{aligned}$$

And the bootstrap analog of the upper bound is

$$\hat{\mathcal{U}}_1^{(b)}(c_1) = \sup_{\gamma_2 \in \mathbb{R}^{p_2}, \gamma_3 \in \mathbb{R}^{p_3}} \tau\Big(\hat{\Sigma}_1^{(b)}, \hat{\zeta}_{2,1}^{(b)}, \ldots, \hat{\zeta}_{2,K_2}^{(b)}, \sqrt{n}(\hat{\mathbb{P}}_n^{(b)} - \mathbb{P}_n), \hat{\mathbb{P}}_n^{(b)}, \widetilde{\mathbb{V}}_{2,n}^{(b)}, \hat{\mathcal{A}}_2^{(b)}, \widetilde{\mathcal{A}}_2^{(b)},$$
$$(\hat{\beta}_1^{\intercal}, \hat{\beta}_2^{\intercal})^{\intercal}, \sqrt{n}(\hat{\beta}_2^{(b)} - \hat{\beta}_2), \sqrt{n}\hat{\beta}_2, \gamma_2, \sqrt{n}\hat{\beta}_3, \gamma_3\Big).$$

Results for $t = 1$ in Theorems 2.1 and 2.2 rely on the limiting behavior of desired quantities (e.g. joint convergence of $(\mathbb{G}_n, \widetilde{\mathbb{V}}_{2,n}, \sqrt{n}(\hat{\beta}_2 - \beta_2^*))$ to $(\mathbb{G}_\infty, \widetilde{\mathbb{V}}_{2,\infty}, \mathbb{V}_{2,\infty})$, where the formula of $\widetilde{\mathbb{V}}_{2,\infty}$ is given in (28)), desired continuity of $w_{1,\cdot}$'s, negligibility of the error terms $\rho_{1,\cdot}$'s, and the fact that $\mathrm{P}(\widetilde{\mathbb{V}}_{2,\infty} \in \Pi_{p_2} C_b(\mathbb{R}^{p_3})) = 1$, where $\widetilde{\mathbb{V}}_{2,\infty}$ is given in (28) and we use $\Pi_{p_2} C_b(\mathbb{R}^{p_3})$ to denote the set of continuous bounded p_2 vector-valued functions on $\mathbb{R}^{p_3}$.

The limiting process of $\mathbb{G}_n$ and limiting distribution of $\sqrt{n}(\hat{\beta}_2 - \beta_2^*))$ have been given in previous sections. The limiting process of $\widetilde{\mathbb{V}}_{2,n}$ is given in Lemma 2.5 below. Joint convergence of these three quantities can be obtained using similar arguments in the proof of Lemma 2.5. The continuity of $w_{1,\cdot}$'s and negligibility of the error terms can be proved using similar arguments as before. To establish $\mathrm{P}(\widetilde{\mathbb{V}}_{2,\infty} \in \Pi_{p_2} C_b(\mathbb{R}^{p_3})) = 1$, we only need to show that the sample paths for each component $\widetilde{\mathbb{V}}_{2,\infty}$ are continuous w.p.1. Note that the j-th component of $\widetilde{\mathbb{V}}_{2,\infty}(\gamma_3)$ can be written as

$$w_{2,1}(\Sigma_{2,\infty}, \mathbb{G}_\infty, P, \mathbb{V}_{3,\infty}, (\beta_2^{*\intercal}, \beta_3^{*\intercal})^{\intercal}) + w_{2,2}(\Sigma_{2,\infty}, P, \mathbb{V}_{3,\infty}, \gamma_3)$$

with c_2 = the j-th column of $I_{p_2 \times p_2}$. The sample path continuity follows by noticing that $w_{2,2}$ is continuous at points $(\Sigma_{2,\infty}, P, \mathbb{R}^{p_3}, \gamma_3)$ uniformly in $\gamma_3 \in \mathbb{R}^{p_3}$. $\square$

Below we give two lemmas that describe the limiting behavior of the desired quantities.

Lemma 2.5. *Assume (A1)-(A4). Then*

$$\widetilde{\mathbb{V}}_{2,n} \rightsquigarrow_P \widetilde{\mathbb{V}}_{2,\infty}, \widetilde{\mathbb{V}}_{2,n}^{P_n} \rightsquigarrow_{P_n} \widetilde{\mathbb{V}}_{2,\infty}, \text{ and } \widetilde{\mathbb{V}}_{2,n}^{(b)} \rightsquigarrow_{P_M} \widetilde{\mathbb{V}}_{2,\infty} \text{ in } P\text{-probability}$$

in $\Pi_{p_2} l^\infty(\mathbb{R}^{p_3})$*, where*

$$\begin{aligned}\widetilde{\mathbb{V}}_{2,\infty}(\gamma_3) = {} & \mathbb{W}'_{2,\infty} + \Sigma_{2,\infty}^{-1} PB_2 \max_{i \in \mathcal{A}_3^*(H_{3,1})} H_{3,1}^\intercal \mathbb{V}_{3,\infty,i} 1_{\#\mathcal{A}_3^*(H_{3,1})=1} \\ & + \Sigma_{2,\infty}^{-1} PB_2 \left(\max_{i \in \mathcal{A}_3^*(H_{3,1})} H_{3,1}^\intercal (\mathbb{V}_{3,\infty,i} + \gamma_{3,i}) - \max_{i \in \mathcal{A}_3^*(H_{3,1})} H_{3,1}^\intercal \gamma_{3,i} \right) 1_{\#\mathcal{A}_3^*(H_{3,1})>1}. \quad (28)\end{aligned}$$

Proof. To prove the first result, we only need to show that

$$\sup_{f \in BL_1(\Pi_{p_2} l^\infty(\mathbb{R}^{p_3}))} \left| \mathbb{E}^* f(\widetilde{\mathbb{V}}_{2,n}) - \mathbb{E} f(\widetilde{\mathbb{V}}_{2,\infty}) \right| \to 0.$$

For convenience, we use $\widetilde{\mathbb{V}}^{[j]}$ to denote the j-th component of $\widetilde{\mathbb{V}}$ for any $\widetilde{\mathbb{V}} \in \Pi_{p_2} l^\infty(\mathbb{R}^{p_3})$. Then

$$\begin{aligned}\sup_{f \in BL_1(\Pi_{p_2} l^\infty(\mathbb{R}^{p_3}))} \left| \mathbb{E}^* f(\widetilde{\mathbb{V}}_{2,n}) - \mathbb{E} f(\widetilde{\mathbb{V}}_{2,\infty}) \right| & \le \sup_{f \in BL_1(\Pi_{p_2} l^\infty(\mathbb{R}^{p_3}))} \mathbb{E}^* \left| f(\widetilde{\mathbb{V}}_{2,n}) - f(\widetilde{\mathbb{V}}_{2,\infty}) \right| \\ & \le \mathbb{E}^* \sup_{j \in \{1,\dots,p_2\}, \gamma_3 \in \mathbb{R}^{p_3}} \left| \widetilde{\mathbb{V}}_{2,n}^{[j]}(\gamma_3) - \widetilde{\mathbb{V}}_{2,\infty}^{[j]}(\gamma_3) \right| \\ & \le \sum_{j=1}^{p_2} \mathbb{E}^* \sup_{\gamma_3 \in \mathbb{R}^{p_3}} \left| \widetilde{\mathbb{V}}_{2,n}^{[j]}(\gamma_3) - \widetilde{\mathbb{V}}_{2,\infty}^{[j]}(\gamma_3) \right|,\end{aligned}$$

where the second inequality follows from the fact that $f \in BL_1(\Pi_{p_2} l^\infty(\mathbb{R}^{p_3}))$. Define $f_j(\widetilde{\mathbb{V}}^{[j]}) \triangleq \sup_{\gamma_3 \in \mathbb{R}^{p_3}} \left| \widetilde{\mathbb{V}}^{[j]}(\gamma_3) - \widetilde{\mathbb{V}}_{2,\infty}^{[j]}(\gamma_3) \right|$ for any $\widetilde{\mathbb{V}}^{[j]} \in l^\infty(\mathbb{R}^{p_3})$, $j = 1, \dots, p_2$. Then $f_j \in BL_1(l^\infty(\mathbb{R}^{p_3}))$ and $f_j(\widetilde{\mathbb{V}}_{2,\infty}^{[j]}) = 0$. Since $\widetilde{\mathbb{V}}_{2,n}^{[j]} \rightsquigarrow_P \widetilde{\mathbb{V}}_{2,\infty}^{[j]}$, we have

$$\mathbb{E}^* \sup_{\gamma_3 \in \mathbb{R}^{p_3}} \left| \widetilde{\mathbb{V}}_{2,n}^{[j]}(\gamma_3) - \widetilde{\mathbb{V}}_{2,\infty}^{[j]}(\gamma_3) \right| = \mathbb{E}^* f_j(\widetilde{\mathbb{V}}_{2,n}^{[j]}) - \mathbb{E} f_j(\widetilde{\mathbb{V}}_{2,\infty}^{[j]}) \to 0$$

for $j = 1, \ldots, p_2$. This completes the proof of $\widetilde{\mathbb{V}}_{2,n} \rightsquigarrow_P \widetilde{\mathbb{V}}_{2,\infty}$. Similar arguments can be used to prove results 2 and 3, and are omitted. $\square$

Lemma 2.6. *Consistency of $\hat{\beta}_t$. Assume (A1)-(A2), then for each t it follows that $\sqrt{n}(\hat{\beta}_t - \beta_t^*) = O_P(1)$ and $\sqrt{n}(\hat{\beta}_t^{(b)} - \hat{\beta}_t) = O_{P_M}(1)$ in P-probability. If, in addition (A4) holds, then $\sqrt{n}(\hat{\beta}_t - \beta_{t,n}^*) = O_{P_n}(1)$.*

Proof. The proof proceeds by backwards induction. The base case follows immediately since $\sqrt{n}(\hat{\beta}_T - \beta_T^*)$ is the usual least squares estimator and hence is asymptotically normal and thus $O_P(1)$. Suppose, as the inductive step, that $\sqrt{n}(\hat{\beta}_{t+1} - \beta_{t+1}^*) = O_P(1)$, the result follows if we can establish that $\sqrt{n}(\hat{\beta}_t - \beta_t^*) = O_P(1)$. Note that $\sqrt{n}(\hat{\beta}_t - \beta_t^*)$ can be decomposed as follows

$$\sqrt{n}(\hat{\beta}_t - \beta_t^*) = \mathbb{W}'_{t,n} + \hat{\Sigma}_t^{-1}\mathbb{P}_n B_t^\intercal \mathbb{U}_{t+1,n}. \tag{29}$$

The proof that $\mathbb{W}'_{t,n}$ is $O_P(1)$ is immediate and omitted. Consider the second term.

$$\begin{aligned}
\big|\big|\mathbb{P}_n B_t^\intercal \mathbb{U}_{t+1,n}\big|\big| &= \left|\left|\sqrt{n}\mathbb{P}_n B_t^\intercal \left(\max_{1\le i\le K_{t+1}} H_{t+1,1}^\intercal \hat{\beta}_{t+1,i} - \max_{1\le i\le K_{t+1}} H_{t+1,1}^\intercal \beta_{t+1,i}^*\right)\right|\right| \\
&\le \mathbb{P}_n ||B_t||\, \sqrt{n}\left|\max_{1\le i\le K_{t+1}} H_{t+1,1}^\intercal \hat{\beta}_{t+1,i} - \max_{1\le i\le K_{t+1}} H_{t+1,1}^\intercal \beta_{t+1,i}^*\right| \\
&\le \mathbb{P}_n ||B_t|| \max_{1\le i\le K_{t+1}} \left|H_{t+1,1}^\intercal \sqrt{n}(\hat{\beta}_{t+1,i} - \beta_{t+1,i}^*)\right| \\
&\le \mathbb{P}_n ||B_t||\, ||H_{t+1,1}|| \max_{1\le i\le K_{t+1}} \big|\big|\sqrt{n}(\hat{\beta}_{t+1,i} - \beta_{t+1,i}^*)\big|\big| \\
&= O_P(1),
\end{aligned}$$

where the last equality follows from the LLN and the induction hypothesis, the series of inequalities follow from repeated use of the Cauchy-Schwartz inequality and the fact that $\big|\max_z f(z) - \max_z g(z)\big| \le \max_z \big|f(z) - g(z)\big|$. This proves the first part of the result. The second and third parts of the result follows from an identical argument. In particular, $\sqrt{n}(\hat{\beta}_T^{(b)} - \hat{\beta}_T)$ converges to the same limiting distribution as $\sqrt{n}(\hat{\beta}_T - \beta_T^*)$ in probability by the bootstrap central limit theorem (see for example Bickel and Freedman 1981) and hence

satisfies the condition stated in the theorem. The same induction argument succeeds with only minor changes in notation. $\square$

Lemma 2.7. *Assume (A1)-(A3). For any* $t = 1, \ldots, T$*, define* $\Delta_t(h_{t,1}) \triangleq \#\big\{\hat{\mathcal{A}}_t(h_{t,1}) \Delta \mathcal{A}_t^*(h_{t,1})\big\}$*. Let* $\epsilon > 0$ *be arbitrary. There exists subset* $B_{t,\epsilon}$ *of* $\mathbb{R}^{p_t/K_t}$ *satisfying* $P(H_{t,1} \in B_{t,\epsilon}) \geq 1 - \epsilon$*, and* $\sup_{h_{t,1} \in B_{t,\epsilon}} \Delta_t(h_{t,1}) = o_P(1)$*.*

Proof. For any fixed arbitrary $\epsilon > 0$, we can choose a sufficiently small $\delta > 0$ so that

$$P\left(0 < \frac{\left|H_{t,1}^\intercal \beta_{t,i}^* - \max_{j \neq i} H_{t,1}^\intercal \beta_{t,j}^*\right|}{||H_{t,1}||} < \delta\right) \leq \epsilon / K_t$$

for $i = 1, \ldots, K_t$, where we have defined $0/0 = 0$ for convenience. Define $B_{t,\epsilon} \triangleq \bigcap_{i=1}^{K_t} B_{t,\epsilon,i}$, where

$$B_{t,\epsilon,i} \triangleq \left\{h_{t,1} : \frac{\left|h_{t,1}^\intercal \beta_{t,i}^* - \max_{j \neq i} h_{t,1}^\intercal \beta_{t,j}^*\right|}{||h_{t,1}||} = 0 \text{ or } \frac{\left|h_{t,1}^\intercal \beta_{t,i}^* - \max_{j \neq i} H_{t,1}^\intercal \beta_{t,j}^*\right|}{||h_{t,1}||} > \delta\right\}.$$

The union bound ensures that $P(H_{t,1} \in B_{t,\epsilon}) \geq 1 - \epsilon$.

To establish the result, we first show that $\sup_{h_{t,1} \in B_{t,\epsilon}} \#(\hat{\mathcal{A}}_t(h_{t,1}) \setminus \mathcal{A}_t^*(h_{t,1})) = o_P(1)$. Note that $\sup_{h_{t,1} \in B_{t,\epsilon}} \#\big(\hat{\mathcal{A}}_t(h_{t,1}) \setminus \mathcal{A}_t^*(h_{t,1})\big)$ can be decomposed as

$$\sup_{h_{t,1} \in B_{t,\epsilon}} \#(\hat{\mathcal{A}}_t(h_{t,1}) \setminus \mathcal{A}_t^*(h_{t,1})) 1_{\min_{j=1,\ldots,K_t} \hat{T}_{t,j}(h_{t,1}) > \lambda_n}$$
$$+ \sup_{h_{t,1} \in B_{t,\epsilon}} \#(\hat{\mathcal{A}}_t(h_{t,1}) \setminus \mathcal{A}_t^*(h_{t,1})) 1_{\min_{j=1,\ldots,K_t} \hat{T}_{t,j}(h_{t,1}) \leq \lambda_n} \quad (30)$$

For any $h_{t,1} \in B_{t,\epsilon}$, let $j^*(h_{t,1})$ denote an element in $\in \mathcal{A}_t^*(h_{t,1})$. Then

$$\begin{aligned}
&\sup_{h_{t,1}\in B_{t,\epsilon}, i\in\{1,\dots,K_t\}\setminus\mathcal{A}_t^*(h_{t,1})} \frac{\sqrt{n}h_{t,1}^T(\hat{\beta}_{t,i}-\hat{\beta}_{t,j^*(h_{t,1})})}{||h_{t,1}||} \\
&= \sup_{h_{t,1}\in B_{t,\epsilon}, i\in\{1,\dots,K_t\}\setminus\mathcal{A}_t^*(h_{t,1})} \frac{h_{t,1}^T(\mathbb{V}_{t,n,i}-\mathbb{V}_{t,n,j^*(h_{t,1})})+\sqrt{n}h_{t,1}^T(\beta_{t,i}^*-\beta_{t,j^*(h_{t,1})}^*)}{||h_{t,1}||} \\
&\qquad \le (2\,||\mathbb{V}_{t,n}||-\sqrt{n}\delta) < 0
\end{aligned}$$

in probability as $n \to \infty$, where the inequality follows from the definition of $B_{t,\epsilon}$. This together with the definition of $\hat{\mathcal{A}}_t(h_{t,1})$ for $\min_{i=1,\dots,K_t}\hat{T}_{t,i}(h_{t,1}) > \lambda_n$ implies that the first term of (30) is $o_P(1)$.

In addition, it is easy to see that for any $h_{t,1} \in B_{t,\epsilon}, i \in \{1,\dots,K_t\} \setminus \mathcal{A}_t^*(h_{t,1})$,

$$\begin{aligned}
\sqrt{\frac{\hat{T}_{t,i}(h_{t,1})}{\lambda_n}} &\ge \frac{\left|h_{t,1}^\intercal(\mathbb{V}_{t,n,i}+\sqrt{n}\beta_{t,i}^*)-\max_{j\neq i}h_{t,1}^\intercal\left(\mathbb{V}_{t,n,j}+\sqrt{n}\beta_{t,j}^*\right)\right|}{\sqrt{\lambda_n\hat{\nu}_{t,i}}||h_{t,1}||} \\
&\ge \frac{\left|\max_{j\neq i}h_{t,1}^\intercal(\mathbb{V}_{t,n,j}+\sqrt{n}(\beta_{t,j}^*-\beta_{t,i}^*))\right|-\left|h_{t,1}^\intercal\mathbb{V}_{t,n,i}\right|}{\sqrt{\lambda_n\max_{i=1,\dots,K_t}\hat{\nu}_{t,i}}||h_{t,1}||},
\end{aligned}$$

where $\hat{\nu}_{t,i}$ denotes the largest eigenvalue of $\hat{\zeta}_{t,i}$. Notice that the numerator on the right hand side of the above display is further bounded below by

$$\sqrt{n}\max_{j\neq i}h_{t,1}^\intercal(\beta_{t,j}^*-\beta_{t,i}^*)-\max_{j\neq i}|h_{t,1}^\intercal\mathbb{V}_{t,n,j}|-\left|h_{t,1}^\intercal\mathbb{V}_{t,n,i}\right| \ge (\sqrt{n}\delta-2||\mathbb{V}_{t,n}||)||h_{t,1}||.$$

Thus

$$\inf_{h_{t,1}\in B_{t,\epsilon}, i\in\{1,\dots,K_t\}\setminus\mathcal{A}_t^*(h_{t,1})}\sqrt{\frac{\hat{T}_{t,i}(h_{t,1})}{\lambda_n}} \ge \frac{\sqrt{n}\delta-2||\mathbb{V}_{t,n}||}{\sqrt{\lambda_n\max_{j=1,\dots,K_t}\hat{\nu}_{t,j}}} > 1$$

in probability as $n \to \infty$. This together with the definition of $\hat{\mathcal{A}}_t(h_{t,1})$ for $\min_{i=1,\dots,K_t}\hat{T}_{t,i}(h_{t,1}) \le \lambda_n$ implies that the second term of (30) is $o_P(1)$.

Next we show that $\sup_{h_{t,1}\in B_{t,\epsilon}}\#(\mathcal{A}_t^*(h_{t,1})\setminus\hat{\mathcal{A}}_t(h_{t,1})) = o_P(1)$. Again, we decompose

$\sup_{h_{t,1}\in B_{t,\epsilon}} \#(\mathcal{A}_t^*(h_{t,1}) \setminus \hat{\mathcal{A}}_t(h_{t,1}))$ as

$$\sup_{h_{t,1}\in B_{t,\epsilon}} \#\big(\mathcal{A}_t^*(h_{t,1}) \setminus \hat{\mathcal{A}}_t(h_{t,1})\big) 1_{\#\mathcal{A}_t^*(h_{t,1})=1} + \sup_{h_{t,1}\in B_{t,\epsilon}} \#\big(\mathcal{A}_t^*(h_{t,1}) \setminus \hat{\mathcal{A}}_t(h_{t,1})\big) 1_{\#\mathcal{A}_t^*(h_{t,1})>1}. \quad (31)$$

For any $h_{t,1} \in B_{t,\epsilon} \cap \{\#\mathcal{A}_t^*(h_{t,1}) = 1\}$, we have

$$\min_{i=1,\dots,K_t} \sqrt{\frac{\hat{T}_{t,i}(h_{t,1})}{\lambda_n}} \geq \min_{i=1,\dots,K_t} \frac{\left| h_{t,1}^\intercal \mathbb{V}_{t,n,j} - \max_{j\neq i} h_{t,1}^\intercal \left(\mathbb{V}_{t,n,j} + \sqrt{n}(\beta_{t,j}^* - \beta_{t,i}^*)\right) \right|}{\sqrt{\lambda_n \hat{\nu}_{t,i}}||h_{t,1}||}.$$

Using similar arguments as above, it is easy to show that the above display is further bounded below by $(\sqrt{n}\eta - 2||\mathbb{V}_{t,n}||)/\sqrt{\lambda_n \max_{j=1,\dots,K_t} \hat{\nu}_{t,j}}$, which does not depend on $h_{t,1}$. Thus

$$\min_{h_{t,1}\in B_{t,\epsilon}\cap\{\#\mathcal{A}_t^*(h_{t,1})=1\}, i=1,\dots,K_t} \sqrt{\frac{\hat{T}_{t,i}(h_{t,1})}{\lambda_n}} \geq \frac{\sqrt{n}\eta - 2||\mathbb{V}_{t,n}||}{\sqrt{\lambda_n \max_{i=1,\dots,K_t} \hat{\nu}_{t,i}}} > 1$$

in probability as $n \to \infty$. Furthermore, let $j^*(h_{t,1}) \triangleq \mathcal{A}_t^*(h_{t,1})$. Then

$$\begin{aligned}
&\min_{h_{t,1}\in B_{t,\epsilon}\cap\{\#\mathcal{A}_t^*(h_{t,1})=1\}} \frac{\sqrt{n}(h_{t,1}^\intercal \hat{\beta}_{t,j^*(h_{t,1})} - \max_{i\neq j^*(h_{t,1})} h_{t,1}^\intercal \hat{\beta}_{t,i})}{||h_{t,1}||} \\
&\geq \min_{h_{t,1}\in B_{t,\epsilon}\cap\{\#\mathcal{A}_t^*(h_{t,1})=1\}} \frac{\sqrt{n}\max_{i\neq j^*(h_{t,1})} h_{t,1}^\intercal(\beta_{t,j^*(h_{t,1})}^* - \beta_{t,i}^*) - 2\max_{i=1,\dots,K_t} |h_{t,1}^\intercal \mathbb{V}_{t,n,i}|}{||h_{t,1}||} \\
&\geq \sqrt{n}\delta - 2||\mathbb{V}_{t,n}|| > 0
\end{aligned}$$

in probability as $n \to \infty$. This implies that the first term of (31) is $o_P(1)$.

For any $h_{t,1}$ such that $\#\mathcal{A}_t^*(h_{t,1}) > 1$, let $j^*(h_{t,1})$ denote an arbitrary element in $\mathcal{A}_t^*(h_{t,1})$. It is easy to verify that

$$\sqrt{\frac{\hat{T}_{t,j^*(h_{t,1})}(h_{t,1})}{\lambda_n}} \leq \frac{||h_{t,1}||\,||\mathbb{V}_{t,n}|| + \left|\max_{i\neq j^*(h_{t,1})} h_{t,1}^\intercal\big(\mathbb{V}_{t,n,i} + \sqrt{n}(\beta_{t,i}^* - \beta_{t,j^*(h_{t,1})}^*)\big)\right|}{\sqrt{\lambda_n \hat{\sigma}_{t,j^*(h_{t,1})}}||h_{t,1}||},$$

where $\hat{\sigma}_{t,i}$ is the smallest eigenvalue of $\hat{\zeta}_{t,i}$. Next note that when $\#\mathcal{A}_t^*(h_{t,1}) > 1$,

$$\frac{\left|\max_{i \neq j^*(h_{t,1})} h_{t,1}^\intercal \left(\mathbb{V}_{t,n,i} + \sqrt{n}(\beta_{t,i}^* - \beta_{t,j^*(h_{t,1})}^*)\right)\right|}{||h_{t,1}||} \leq \max_{i \neq j^*(h_{t,1})} ||\mathbb{V}_{t,n,i}|| \leq ||\mathbb{V}_{t,n}||.$$

To see this, let $\tilde{i}(h_{t,1}) \triangleq \arg\max_{i \neq j^*(h_{t,1})} h_{t,1}^\intercal \left(\mathbb{V}_{t,n,i} + \sqrt{n}(\beta_{t,i}^* - \beta_{t,j^*(h_{t,1})}^*)\right)$ and $j'(h_{t,1})$ be an element in $\mathcal{A}_t^*(h_{t,1}) \setminus \{j^*(h_{t,1})\}$, and notice that

$$h_{t,1}^\intercal \mathbb{V}_{t,n,j'(h_{t,1})} \leq h_{t,1}^\intercal \left(\mathbb{V}_{t,n,\tilde{i}(h_{t,1})} + \sqrt{n}(\beta_{t,\tilde{i}(h_{t,1})}^* - \beta_{t,j^*(h_{t,1})}^*)\right) \leq h_{t,1}^\intercal \mathbb{V}_{t,n,\tilde{i}(h_{t,1})},$$

where the second inequality makes use of the fact that $h_{t,1}^\intercal(\beta_{t,i}^* - \beta_{t,j^*(h_{t,1})}^*) \leq 0$ with equality holding when $i \in \mathcal{A}_t^*(h_{t,1})$. This result, the preceding discussion, and some algebra show that $\hat{T}_{t,i}(h_{t,1})/\lambda_n$ is bounded above by

$$\sup_{h_{t,1} \in \{h_{t,1} : \#\mathcal{A}_t^*(h_{t,1}) > 1\}, i \in \mathcal{A}_t^*(h_{t,1})} \sqrt{\frac{\hat{T}_{t,i}(h_{t,1})}{\lambda_n}} \leq \frac{2||\mathbb{V}_{t,n}||}{\sqrt{\lambda_n} \min_{i=1,\ldots,K_t} \hat{\sigma}_{t,i}},$$

which is $o_P(1)$. Thus the second term of (31) is $o_P(1)$. This completes the proof. □

3 Bias reduction for non-regular problems

In this section we briefly discuss the issue of bias reduction for non-regular problems. It is now well known that unbiased estimators do not exist for non-smooth functionals (see Robins 2004, appendix I; and Porter and Hirano 2009). Furthermore, it has been shown that attempting to reduce the bias at a non-regular point in the parameter space can dramatically inflate the variance and subsequently the MSE elsewhere in the paramter space (Doss and Sethuraman 1989; Brown and Liu 1993; Chen 2004). Here, we attempt to illustrate this phenomenon in a toy example that is relevant for medical decision making.

Suppose that X_1 and X_2 are independent normal random variables with means μ_1 and

μ_2 respectively, both are assumed to have unit variance. We consider the task of estimating $\theta \triangleq \max(\mu_1, \mu_2)$ based on a single observation $X_1 = x_1$ and $X_2 = x_2$. Notice that this problem corresponds to a toy decision making problem where μ_i denotes the mean response for patients following treatment i. The MLE is given by $\hat{\theta}_{\mathrm{mle}} \triangleq \max(X_1, X_2)$. It is clear that the MLE suffers from upward bias since

$$\theta \triangleq \max(\mu_1, \mu_2) = \max(\mathbb{E}X_1, \mathbb{E}X_2) \leq \mathbb{E}\max(X_1, X_2).$$

It will be convenient to write the $\hat{\theta}_{\mathrm{mle}}$ as

$$\hat{\theta}_{\mathrm{mle}} \triangleq (X_1 + X_2)/2 + |X_1 - X_2|/2.$$

The first term on the right hand side of the above display is the UMVU estimator of θ when there is no treatment effect (e.g. $\mu_1 = \mu_2$). The second term can be seen as an estimator of the advantage of recommending treatment via the decision rule $\arg\max_{i=1,2} X_i$ compared with randomly assigning treatment according to an even odds coin flip. The thresholding estimators of Chakraborty et al. (2009) and Moodie and Richardson (2007) shrink the term $|X_1 - X_2|/2$ towards zero in an attempt to alleviate some of the bias inherent to $\hat{\theta}_{\mathrm{mle}}$. In particular, an analogue of the soft-thresholding estimator of Chakraborty et al. (2009) for this problem is given by

$$\hat{\theta}_{\mathrm{soft}} \triangleq (X_1 + X_2)/2 + \left[1 - \frac{\lambda}{|X_1 - X_2|}\right]_+ |X_1 - X_2|/2$$

where λ denotes a tuning parameter. An analogue of the hard-thesholding estimator of Moodie and Richardson (2007) is given by

$$\hat{\theta}_{\mathrm{hard}} \triangleq (X_1 + X_2)/2 + 1_{|X_1 - X_2| \geq \lambda}|X_1 - X_2|/2,$$

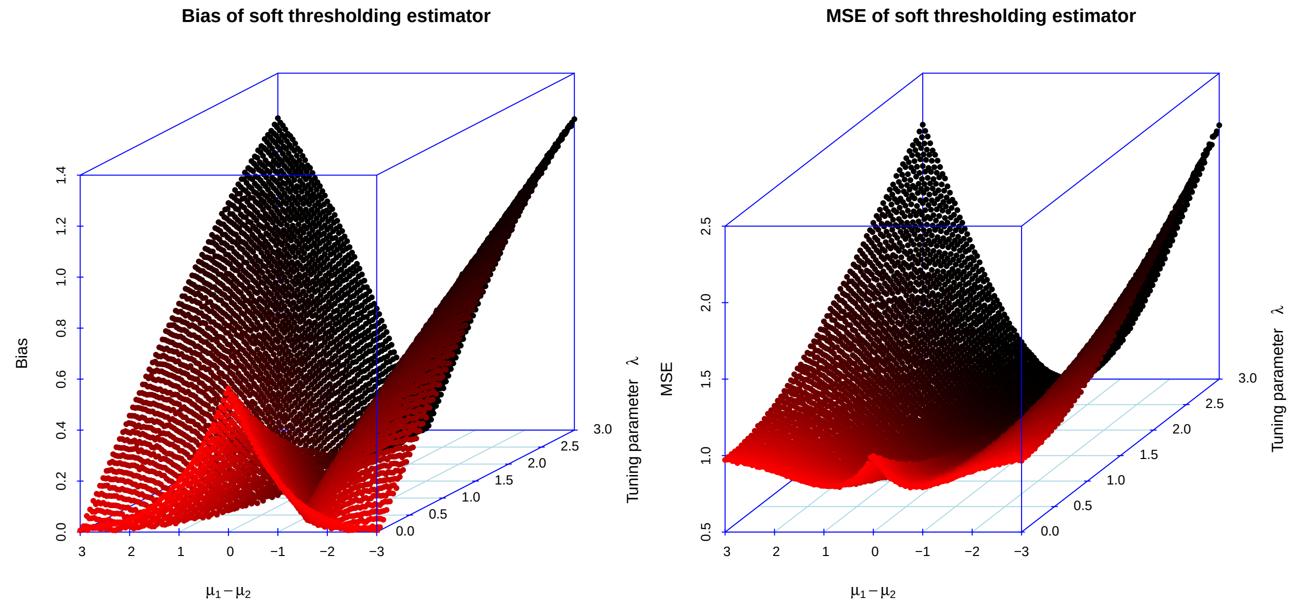

Figure 1: Left: The bias of $\hat{\theta}_{\text{soft}}$ as a function of effect size $\mu_1 - \mu_2$ and tuning parameter λ. Reducing the bias at $\mu_1 - \mu_2 = 0$ requires increasing λ which is seen to dramatically inflate bias elsewhere. Right: The MSE of $\hat{\theta}_{\text{soft}}$ as a function of effect size $\mu_1 - \mu_2$ and tuning parameter λ. Attempting to reduce the bias at $\mu_1 - \mu_2 = 0$ results in a modest reduction in MSE at $\mu_1 - \mu_2 = 0$ but inflates the MSE significantly elsewhere.

again where λ is a tuning parameter. Notice that both estimators reduce to $\hat{\theta}_{\text{mle}}$ when $\lambda = 0$. As we will see, the bias $\hat{\theta}_{\text{mle}}$ is largest when $\mu_1 = \mu_2$. Both $\hat{\theta}_{\text{soft}}$ and $\hat{\theta}_{\text{hard}}$ seek to alleviate some of this bias by shrinking $\hat{\theta}_{\text{mle}}$ towards $(X_1 + X_2)/2$ whenever $|X_1 - X_2|$ is small.

Figure (3) shows the bias and MSE of the soft-threshold estimator $\hat{\theta}_{\text{soft}}$ as a function of effect size $\mu_1 - \mu_2$ and tuning parameter λ. The figure shows that by increasing λ the bias at $\mu_1 - \mu_2 = 0$ decreases, however, modest increases in λ lead to dramatic increases in bias non-zero values of $\mu_1 - \mu_2$ and subsequently inflate the MSE. Figure (3) shows results of a similar nature for the hard-thresholding estimator $\hat{\theta}_{\text{hard}}$. These figures show that the price of bias reduction at $\mu_1 - \mu_2 = 0$ can be quite severe unless one has very strong prior knowledge about the true value of $\mu_1 - \mu_2$.

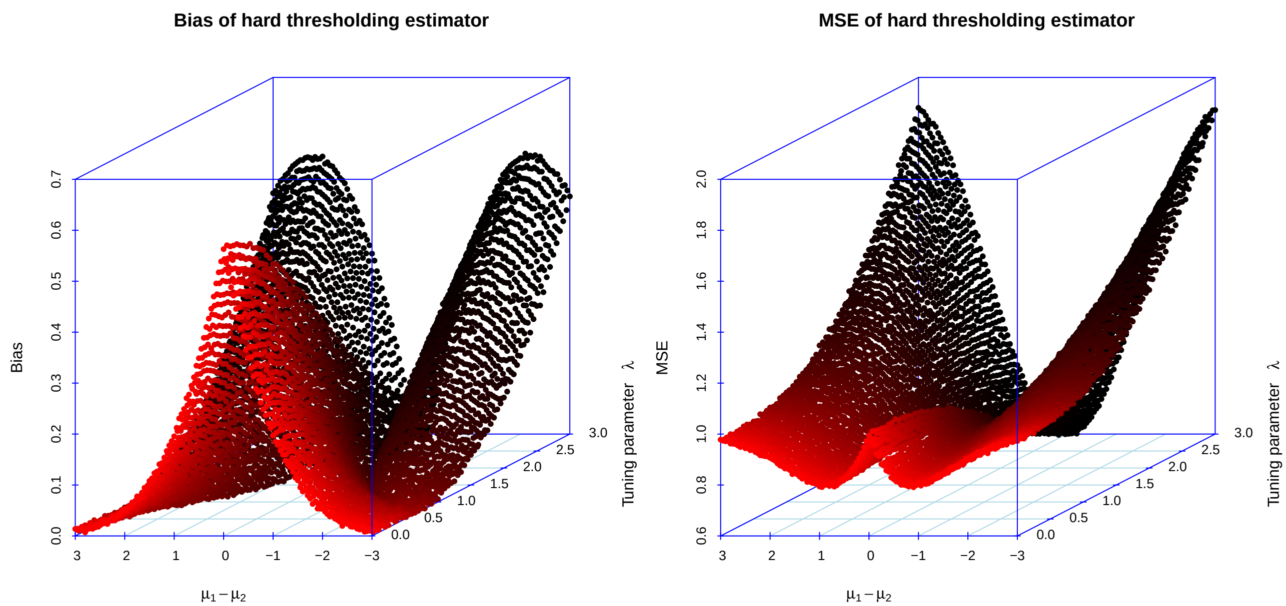

Figure 2: Left: The bias of $\hat{\theta}_{\text{hard}}$ as a function of effect size $\mu_1 - \mu_2$ and tuning parameter λ. Reducing the bias at $\mu_1 - \mu_2 = 0$ requires increasing λ which is seen to dramatically inflate bias elsewhere. Right: The MSE of $\hat{\theta}_{\text{hard}}$ as a function of effect size $\mu_1 - \mu_2$ and tuning parameter λ. Attempting to reduce the bias at $\mu_1 - \mu_2 = 0$ results in a modest reduction in MSE at $\mu_1 - \mu_2 = 0$ but inflates the MSE significantly elsewhere.

4 Additional empirical results

Here, we present additional empirical results for the ACI and competitors. We give results for the generative models in the main body of the paper with varying dataset sizes, for generative models with three treatments at the second stage, and for generative models with three stages of binary treatments. All of the results in this section are based on 1000 Monte Carlo repetitions, and for the ACI we use the tuning parameter $\lambda_n = \log\log n$.

4.1 Varying dataset size

First, we present a suite of experiments with the two-stage, two-action models presented in the main body of the paper, with varying data set size N. Tables 1 through 12 show our results.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.892*	0.908*	0.924*	0.925*	0.940	0.930*	0.936	0.925*	0.931*
PPE	0.926*	0.930*	0.933*	0.934*	0.934*	0.907*	0.928*	0.910*	0.909*
ST	0.935*	0.930*	0.889*	0.878*	0.891*	0.620*	0.687*	0.686*	0.663*
ACI	0.956	0.964	0.954	0.955	0.950	0.957	0.948	0.956	0.957

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.899*	0.915*	0.947	0.949	0.939	0.967	0.961	0.946	0.949
PPE	0.949	0.946	0.952	0.948	0.941	0.948	0.958	0.949	0.949
ST	0.952	0.945	0.935*	0.929*	0.935*	0.644*	0.780*	0.869*	0.851*
ACI	0.970	0.976	0.969	0.970	0.956	0.973	0.965	0.972	0.975

$N = 500$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.892*	0.906*	0.935*	0.933*	0.929*	0.942	0.943	0.931*	0.934*
PPE	0.936	0.938	0.941	0.937	0.929*	0.934*	0.938	0.943	0.937
ST	0.956	0.949	0.923*	0.917*	0.910*	0.664*	0.790*	0.895*	0.875*
ACI	0.965	0.976	0.964	0.968	0.952	0.950	0.944	0.966	0.967

$N = 1000$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.907*	0.933*	0.933*	0.943	0.944	0.945	0.951	0.936	0.940
PPE	0.949	0.938	0.949	0.947	0.952	0.942	0.949	0.942	0.944
ST	0.953	0.933*	0.944	0.934*	0.934*	0.813*	0.880*	0.921*	0.892*
ACI	0.968	0.980	0.968	0.971	0.961	0.946	0.951	0.968	0.971

Table 1: Monte Carlo estimates of coverage probability of confidence intervals for $\beta^*_{1,0,1}$ (intercept term) at the 95% nominal level. Generative models have two stages and two actions per stage. Estimates are constructed using 1000 datasets of size 150, 300, 500, and 1000 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.404*	0.404*	0.430*	0.429*	0.457	0.449*	0.450	0.428*	0.428*
PPE	0.376*	0.376*	0.418*	0.418*	0.451*	0.448*	0.453*	0.410*	0.410*
ST	0.344*	0.344*	0.427*	0.427*	0.466*	0.469*	0.474*	0.430*	0.428*
ACI	0.518	0.518	0.487	0.487	0.486	0.494	0.476	0.497	0.498

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.284*	0.284*	0.300	0.300	0.320	0.314	0.314	0.299	0.299
PPE	0.264	0.264	0.292	0.292	0.316	0.316	0.317	0.292	0.292
ST	0.240	0.240	0.289*	0.289*	0.319*	0.326*	0.324*	0.307*	0.307*
ACI	0.367	0.367	0.343	0.343	0.341	0.338	0.328	0.343	0.344

$N = 500$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.218*	0.218*	0.232*	0.232*	0.248*	0.243	0.243	0.232*	0.232*
PPE	0.203	0.203	0.226	0.226	0.245*	0.247*	0.245	0.226	0.226
ST	0.184	0.185	0.221*	0.222*	0.245*	0.253*	0.251*	0.232*	0.232*
ACI	0.284	0.284	0.265	0.265	0.265	0.255	0.249	0.265	0.265

$N = 1000$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.155*	0.155*	0.164*	0.164	0.175	0.171	0.171	0.164	0.164
PPE	0.144	0.144	0.159	0.160	0.173	0.173	0.172	0.159	0.159
ST	0.131	0.131*	0.156	0.156*	0.172*	0.179*	0.176*	0.159*	0.159*
ACI	0.202	0.202	0.188	0.188	0.187	0.174	0.172	0.188	0.188

Table 2: Monte Carlo estimates of the mean width of confidence intervals for $\beta^*_{1,0,1}$ (intercept term) at the 95% nominal level. Generative models have two stages and two actions per stage. Estimates are constructed using 1000 datasets of size 150, 300, 500, and 1000 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.942	0.944	0.948	0.948	0.928*	0.942	0.939	0.944	0.944
PPE	0.946	0.946	0.945	0.945	0.931*	0.936	0.939	0.946	0.947
ST	0.946	0.946	0.950	0.950	0.941	0.941	0.941	0.945	0.946
ACI	0.964	0.966	0.958	0.957	0.941	0.947	0.940	0.954	0.954

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.942	0.947	0.952	0.950	0.948	0.946	0.958	0.945	0.945
PPE	0.944	0.946	0.953	0.953	0.943	0.942	0.956	0.945	0.945
ST	0.945	0.945	0.948	0.949	0.951	0.940	0.955	0.944	0.944
ACI	0.960	0.959	0.957	0.957	0.955	0.946	0.958	0.951	0.951

$N = 500$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.948	0.951	0.954	0.953	0.948	0.952	0.953	0.953	0.953
PPE	0.948	0.950	0.955	0.953	0.951	0.951	0.952	0.949	0.948
ST	0.948	0.948	0.954	0.953	0.951	0.952	0.949	0.952	0.951
ACI	0.967	0.966	0.964	0.964	0.961	0.952	0.953	0.959	0.959

$N = 1000$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.941	0.945	0.938	0.944	0.937	0.941	0.941	0.943	0.942
PPE	0.942	0.944	0.939	0.942	0.936	0.940	0.941	0.943	0.943
ST	0.945	0.947	0.944	0.943	0.941	0.939	0.945	0.942	0.943
ACI	0.963	0.961	0.955	0.955	0.945	0.941	0.941	0.947	0.947

Table 3: Monte Carlo estimates of coverage probability of confidence intervals for $\beta^*_{1,0,2}$ (main effect of history) at the 95% nominal level. Generative models have two stages and two actions per stage. Estimates are constructed using 1000 datasets of size 150, 300, 500, and 1000 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.331	0.331	0.333	0.333	0.379*	0.354	0.355	0.329	0.329
PPE	0.330	0.330	0.332	0.332	0.376*	0.350	0.353	0.329	0.329
ST	0.328	0.328	0.332	0.332	0.384	0.360	0.361	0.329	0.329
ACI	0.360	0.360	0.347	0.348	0.392	0.359	0.358	0.339	0.339

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.231	0.231	0.232	0.232	0.265	0.246	0.246	0.229	0.229
PPE	0.230	0.230	0.231	0.231	0.263	0.245	0.246	0.229	0.229
ST	0.229	0.229	0.231	0.231	0.266	0.250	0.249	0.229	0.229
ACI	0.251	0.251	0.242	0.242	0.275	0.248	0.247	0.233	0.233

$N = 500$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.178	0.178	0.179	0.179	0.205	0.190	0.190	0.177	0.177
PPE	0.178	0.178	0.178	0.178	0.204	0.190	0.190	0.177	0.177
ST	0.177	0.177	0.178	0.178	0.205	0.193	0.192	0.177	0.177
ACI	0.194	0.194	0.187	0.187	0.213	0.191	0.191	0.179	0.179

$N = 1000$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.126	0.126	0.126	0.126	0.145	0.134	0.134	0.124	0.124
PPE	0.125	0.125	0.126	0.126	0.144	0.134	0.134	0.124	0.124
ST	0.124	0.124	0.125	0.125	0.144	0.135	0.135	0.124	0.124
ACI	0.137	0.137	0.132	0.132	0.150	0.134	0.134	0.126	0.126

Table 4: Monte Carlo estimates of the mean width of confidence intervals for $\beta^*_{1,0,2}$ (main effect of history) at the 95% nominal level. Generative models have two stages and two actions per stage. Estimates are constructed using 1000 datasets of size 150, 300, 500, and 1000 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.934*	0.935*	0.930*	0.933*	0.938	0.928*	0.939	0.925*	0.928*
PPE	0.931*	0.940	0.938	0.940	0.946	0.912*	0.931*	0.904*	0.903*
ST	0.948	0.945	0.938	0.942	0.952	0.943	0.919*	0.759*	0.762*
ACI	0.992	0.992	0.968	0.972	0.957	0.955	0.950	0.964	0.965

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.952	0.952	0.948	0.952	0.943	0.936	0.941	0.949	0.951
PPE	0.951	0.952	0.960	0.959	0.956	0.907*	0.944	0.952	0.954
ST	0.951	0.949	0.938	0.941	0.949	0.951	0.920*	0.877*	0.883*
ACI	0.994	0.994	0.975	0.976	0.962	0.957	0.950	0.977	0.976

$N = 500$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.947	0.944	0.947	0.947	0.943	0.946	0.944	0.943	0.946
PPE	0.952	0.945	0.950	0.951	0.940	0.919*	0.945	0.945	0.944
ST	0.965	0.965	0.953	0.959	0.951	0.927*	0.910*	0.924*	0.935*
ACI	0.992	0.992	0.976	0.980	0.956	0.958	0.947	0.975	0.978

$N = 1000$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.948	0.949	0.934*	0.939	0.950	0.954	0.951	0.939	0.947
PPE	0.948	0.949	0.948	0.945	0.952	0.941	0.948	0.950	0.950
ST	0.956	0.955	0.959	0.955	0.954	0.935*	0.924*	0.947	0.958
ACI	0.998	0.995	0.972	0.973	0.963	0.954	0.951	0.972	0.977

Table 5: Monte Carlo estimates of coverage probability of confidence intervals for $\beta^*_{1,1,1}$ (main effect of treatment) at the 95% nominal level. Generative models have two stages and two actions per stage. Estimates are constructed using 1000 datasets of size 150, 300, 500, and 1000 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.385*	0.385*	0.430*	0.430*	0.457	0.436*	0.451	0.428*	0.428*
PPE	0.365*	0.366	0.419	0.419	0.452	0.418*	0.452*	0.404*	0.403*
ST	0.339	0.339	0.426	0.427	0.469	0.436	0.480*	0.426*	0.424*
ACI	0.502	0.502	0.488	0.488	0.487	0.475	0.477	0.491	0.491

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.269	0.269	0.300	0.300	0.320	0.309	0.313	0.299	0.299
PPE	0.256	0.256	0.292	0.292	0.316	0.297*	0.317	0.290	0.290
ST	0.237	0.237	0.289	0.289	0.320	0.313	0.327*	0.306*	0.307*
ACI	0.354	0.354	0.342	0.342	0.341	0.327	0.327	0.342	0.342

$N = 500$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.208	0.208	0.232	0.232	0.248	0.242	0.244	0.232	0.232
PPE	0.197	0.197	0.226	0.226	0.245	0.234*	0.245	0.226	0.226
ST	0.182	0.183	0.222	0.222	0.246	0.252*	0.253*	0.232*	0.233*
ACI	0.275	0.275	0.265	0.265	0.265	0.250	0.250	0.265	0.265

$N = 1000$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.147	0.147	0.164*	0.164	0.175	0.171	0.171	0.164	0.164
PPE	0.139	0.139	0.160	0.160	0.173	0.170	0.172	0.160	0.160
ST	0.129	0.129	0.156	0.156	0.172	0.184*	0.177*	0.159	0.159
ACI	0.195	0.195	0.188	0.188	0.187	0.173	0.173	0.188	0.188

Table 6: Monte Carlo estimates of the mean width of confidence intervals for $\beta^*_{1,1,1}$ (main effect of treatment) at the 95% nominal level. Generative models have two stages and two actions per stage. Estimates are constructed using 1000 datasets of size 150, 300, 500, and 1000 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.944	0.943	0.948	0.949	0.944	0.952	0.956	0.947	0.947
PPE	0.944	0.945	0.951	0.951	0.941	0.947	0.954	0.949	0.948
ST	0.946	0.946	0.950	0.949	0.955	0.950	0.952	0.948	0.948
ACI	0.963	0.963	0.959	0.959	0.955	0.953	0.957	0.955	0.955

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.960	0.958	0.953	0.954	0.949	0.950	0.948	0.961	0.961
PPE	0.957	0.956	0.954	0.954	0.945	0.945	0.948	0.961	0.961
ST	0.954	0.954	0.952	0.951	0.943	0.946	0.949	0.957	0.956
ACI	0.975	0.975	0.964	0.963	0.955	0.951	0.949	0.964	0.964

$N = 500$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.953	0.952	0.945	0.947	0.936	0.951	0.941	0.946	0.947
PPE	0.953	0.954	0.944	0.945	0.938	0.951	0.941	0.946	0.946
ST	0.947	0.947	0.945	0.945	0.938	0.945	0.938	0.945	0.944
ACI	0.966	0.966	0.956	0.956	0.948	0.952	0.941	0.952	0.951

$N = 1000$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.954	0.955	0.951	0.955	0.956	0.957	0.953	0.953	0.953
PPE	0.954	0.955	0.952	0.954	0.959	0.957	0.953	0.952	0.952
ST	0.953	0.953	0.951	0.952	0.954	0.959	0.954	0.953	0.953
ACI	0.967	0.969	0.959	0.961	0.965	0.958	0.953	0.954	0.954

Table 7: Monte Carlo estimates of coverage probability of confidence intervals for $\beta^*_{1,1,2}$ (interaction between history and treatment) at the 95% nominal level. Generative models have two stages and two actions per stage. Estimates are constructed using 1000 datasets of size 150, 300, 500, and 1000 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.331	0.331	0.333	0.332	0.363	0.354	0.355	0.329	0.329
PPE	0.330	0.330	0.332	0.332	0.361	0.350	0.353	0.328	0.328
ST	0.328	0.328	0.332	0.332	0.366	0.359	0.360	0.329	0.329
ACI	0.360	0.360	0.347	0.347	0.378	0.359	0.358	0.339	0.339

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.231	0.231	0.231	0.231	0.254	0.246	0.246	0.228	0.228
PPE	0.230	0.230	0.231	0.231	0.252	0.244	0.246	0.228	0.228
ST	0.228	0.228	0.230	0.230	0.254	0.250	0.249	0.229	0.229
ACI	0.251	0.250	0.241	0.241	0.264	0.248	0.247	0.233	0.233

$N = 500$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.178	0.178	0.178	0.178	0.197	0.190	0.190	0.176	0.176
PPE	0.177	0.177	0.178	0.178	0.196	0.189	0.190	0.176	0.176
ST	0.176	0.176	0.178	0.178	0.196	0.193	0.191	0.176	0.176
ACI	0.194	0.194	0.186	0.186	0.205	0.191	0.191	0.179	0.179

$N = 1000$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.126	0.126	0.126	0.126	0.139	0.134	0.134	0.124	0.124
PPE	0.125	0.125	0.126	0.126	0.138	0.134	0.134	0.124	0.124
ST	0.125	0.125	0.125	0.125	0.139	0.135	0.134	0.124	0.124
ACI	0.137	0.137	0.132	0.132	0.145	0.134	0.134	0.126	0.126

Table 8: Monte Carlo estimates of the mean width of confidence intervals for $\beta^*_{1,1,2}$ (interaction between history and treatment) at the 95% nominal level. Generative models have two stages and two actions per stage. Estimates are constructed using 1000 datasets of size 150, 300, 500, and 1000 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.942	0.941	0.940	0.942	0.943	0.929*	0.941	0.940	0.939
PPE	0.941	0.938	0.945	0.943	0.937	0.917*	0.935*	0.928*	0.928*
ST	0.943	0.943	0.932*	0.934*	0.940	0.934*	0.928*	0.821*	0.825*
ACI	0.984	0.985	0.963	0.964	0.948	0.946	0.949	0.964	0.967

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.953	0.953	0.954	0.959	0.955	0.953	0.961	0.954	0.961
PPE	0.958	0.955	0.965	0.966	0.956	0.937	0.959	0.962	0.965
ST	0.956	0.958	0.945	0.950	0.958	0.956	0.934*	0.901*	0.906*
ACI	0.988	0.989	0.977	0.979	0.961	0.966	0.966	0.979	0.980

$N = 500$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.951	0.951	0.948	0.950	0.952	0.954	0.954	0.943	0.944
PPE	0.952	0.951	0.953	0.952	0.953	0.938	0.953	0.950	0.949
ST	0.950	0.950	0.956	0.957	0.950	0.938	0.931*	0.933*	0.940
ACI	0.988	0.988	0.968	0.973	0.958	0.957	0.954	0.965	0.971

$N = 1000$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.954	0.949	0.945	0.946	0.953	0.948	0.949	0.956	0.951
PPE	0.958	0.958	0.952	0.951	0.955	0.942	0.948	0.953	0.950
ST	0.966	0.965	0.955	0.960	0.956	0.938	0.934*	0.954	0.954
ACI	0.992	0.991	0.972	0.978	0.962	0.951	0.950	0.972	0.978

Table 9: Monte Carlo estimates of coverage probability of confidence intervals for the contrast $\beta^*_{1,1,1}+\beta^*_{1,1,2}$ (effect of action for history = 1) at the 95% nominal level. Generative models have two stages and two actions per stage. Estimates are constructed using 1000 datasets of size 150, 300, 500, and 1000 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.506	0.506	0.542	0.542	0.580	0.559*	0.571	0.538	0.538
PPE	0.491	0.491	0.533	0.533	0.578	0.544*	0.571*	0.520*	0.519*
ST	0.471	0.471	0.539*	0.539*	0.600	0.563*	0.598*	0.537*	0.536*
ACI	0.622	0.622	0.600	0.600	0.596	0.598	0.595	0.602	0.603

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.355	0.355	0.378	0.378	0.406	0.394	0.397	0.377	0.377
PPE	0.344	0.344	0.372	0.372	0.404	0.383	0.400	0.369	0.369
ST	0.329	0.329	0.369	0.369	0.412	0.399	0.410*	0.383*	0.383*
ACI	0.439	0.439	0.421	0.421	0.417	0.411	0.409	0.420	0.420

$N = 500$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.273	0.274	0.293	0.293	0.315	0.307	0.308	0.291	0.292
PPE	0.265	0.265	0.288	0.288	0.314	0.301	0.310	0.287	0.287
ST	0.254	0.254	0.284	0.284	0.318	0.317	0.315*	0.292*	0.292
ACI	0.340	0.340	0.327	0.327	0.324	0.315	0.314	0.325	0.325

$N = 1000$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.193	0.194	0.207	0.207	0.222	0.217	0.217	0.206	0.206
PPE	0.187	0.187	0.203	0.203	0.222	0.216	0.217	0.202	0.202
ST	0.180	0.180	0.200	0.200	0.223	0.228	0.221*	0.202	0.202
ACI	0.242	0.242	0.231	0.231	0.229	0.218	0.218	0.230	0.230

Table 10: Monte Carlo estimates of the mean width of confidence intervals for the contrast $\beta^*_{1,1,1}+\beta^*_{1,1,2}$ (effect of action for history = 1) at the 95% nominal level. Generative models have two stages and two actions per stage. Estimates are constructed using 1000 datasets of size 150, 300, 500, and 1000 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.958	0.954	0.944	0.945	0.935*	0.943	0.949	0.939	0.940
PPE	0.954	0.952	0.950	0.950	0.938	0.935*	0.941	0.925*	0.926*
ST	0.964	0.964	0.940	0.943	0.938	0.951	0.929*	0.829*	0.832*
ACI	0.985	0.985	0.970	0.972	0.960	0.960	0.953	0.966	0.969

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.949	0.951	0.941	0.942	0.947	0.936	0.945	0.939	0.939
PPE	0.949	0.952	0.942	0.943	0.946	0.920*	0.945	0.943	0.948
ST	0.950	0.950	0.937	0.942	0.943	0.940	0.923*	0.904*	0.908*
ACI	0.984	0.985	0.964	0.966	0.965	0.951	0.949	0.968	0.971

$N = 500$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.943	0.944	0.934*	0.942	0.933*	0.944	0.944	0.938	0.946
PPE	0.949	0.947	0.947	0.950	0.942	0.927*	0.946	0.946	0.947
ST	0.963	0.961	0.943	0.946	0.937	0.929*	0.920*	0.917*	0.922*
ACI	0.986	0.985	0.962	0.967	0.953	0.950	0.948	0.961	0.966

$N = 1000$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.947	0.940	0.937	0.944	0.944	0.950	0.946	0.948	0.948
PPE	0.947	0.946	0.955	0.951	0.944	0.948	0.946	0.953	0.954
ST	0.948	0.946	0.961	0.961	0.953	0.936	0.929*	0.954	0.961
ACI	0.989	0.990	0.971	0.974	0.962	0.953	0.946	0.971	0.977

Table 11: Monte Carlo estimates of coverage probability of confidence intervals for the contrast $\beta^*_{1,1,1} - \beta^*_{1,1,2}$ (effect of action for history = -1) at the 95% nominal level. Generative models have two stages and two actions per stage. Estimates are constructed using 1000 datasets of size 150, 300, 500, and 1000 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.508	0.508	0.544	0.544	0.587*	0.563	0.575	0.539	0.538
PPE	0.491	0.491	0.534	0.534	0.577	0.545*	0.573	0.520*	0.519*
ST	0.471	0.471	0.541	0.542	0.588	0.566	0.601*	0.538*	0.537*
ACI	0.624	0.624	0.601	0.601	0.630	0.600	0.600	0.604	0.604

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.354	0.354	0.379	0.378	0.410	0.398	0.400	0.376	0.376
PPE	0.343	0.343	0.372	0.372	0.404	0.385*	0.402	0.369	0.369
ST	0.329	0.329	0.369	0.370	0.404	0.403	0.412*	0.382*	0.382*
ACI	0.439	0.439	0.420	0.420	0.442	0.415	0.413	0.419	0.419

$N = 500$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.273	0.273	0.293*	0.293	0.318*	0.309	0.310	0.291	0.291
PPE	0.265	0.265	0.288	0.288	0.313	0.302*	0.311	0.286	0.286
ST	0.253	0.253	0.284	0.284	0.311	0.318*	0.318*	0.292*	0.292*
ACI	0.340	0.340	0.325	0.325	0.343	0.316	0.315	0.324	0.324

$N = 1000$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.193	0.193	0.206	0.206	0.224	0.218	0.218	0.205	0.205
PPE	0.187	0.187	0.203	0.203	0.221	0.217	0.218	0.202	0.202
ST	0.179	0.179	0.200	0.200	0.219	0.229	0.223*	0.201	0.201
ACI	0.241	0.241	0.230	0.230	0.242	0.219	0.219	0.229	0.229

Table 12: Monte Carlo estimates of the mean width of confidence intervals for the contrast $\beta^*_{1,1,1} - \beta^*_{1,1,2}$ (effect of action for history = -1) at the 95% nominal level. Generative models have two stages and two actions per stage. Estimates are constructed using 1000 datasets of size 150, 300, 500, and 1000 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

4.2 Models with ternary actions

Here, we present results using a suite of examples similar to those of Chakraborty et al. (2009), but that have three possible treatments at the second stage. These models are defined as follows:

- $X_i \in \{-1, 1\}$ for $i \in \{1, 2\}$, $A_1 \in \{-1, 1\}$, and $A_2 \in \{(0, -0.5)^\intercal, (-1, 0.5)^\intercal, (1, 0.5)^\intercal\}$
- $P(A_1 = 1) = P(A_1 = -1) = 1/2$,
  $P(A_2 = (0, -1)^\intercal) = P(A_2 = (-1, 0.5)^\intercal) = P(A_2 = (1, 0.5)^\intercal) = 1/3$
- $P(X_1 = 1) = P(X_1 = -1) = 1/2$, $P(X_2 = 1|X_1, A_1) = \text{expit}(\delta_1 X_1 + \delta_2 A_1)$
- $Y_1 \triangleq 0$,
  $Y_2 = \xi_1 + \xi_2 X_1 + \xi_3 A_1 + \xi_4 X_1 A_1 + (\xi_5, \xi_6) A_2 + X_2(\xi_7, \xi_8) A_2 + A_1(\xi_9, \xi_{10}) A_2 + \epsilon$, $\epsilon \sim N(0, 1)$

where $\text{expit}(x) = \mathrm{e}^x/(1+\mathrm{e}^x)$. This class is parameterized by twelve values $\xi_1, \xi_2, ..., \xi_{10}, \delta_1, \delta_2$. The analysis model uses histories defined by:

$$
\begin{aligned}
H_{2,0} &= (1, X_1, A_1, X_1 A_1, X_2)^\intercal && (32) \\
H_{2,1} &= (1, X_2, A_1)^\intercal && (33) \\
H_{1,0} &= (1, X_1)^\intercal && (34) \\
H_{1,1} &= (1, X_1)^\intercal. && (35)
\end{aligned}
$$

Our working models are given by $Q_2(H_2, A_2; \beta_2) \triangleq H_{2,0}^\intercal \beta_{2,0} + H_{2,1}^\intercal \beta_{2,1,1} A_{2,1} + H_{2,1}^\intercal \beta_{2,1,2} A_{2,2}$ and $Q_1(H_1, A_1; \beta_1) \triangleq H_{1,0}^\intercal \beta_{1,0} + H_{1,1}^\intercal \beta_{1,1} A_1$. In Table 4.2, for each of these models we give the probability p of generating a history where each of the three possible treatments at the second stage have exactly the same effect. This is analogous to having the second stage action show no effect in a binary model. Furthermore, because of the Helmert encoding we have used in our analysis models, and because of the structure of ξ, it happens that the

Example	ξ	δ	Regularity
1	$(0,0,0,0,0,0,0,0,0,0)^\intercal$	$(0.5,0.5)^\intercal$	$p=1, \phi=0/0$
2	$(0,0,0,0,0.01,0.01,0,0,0,0)^\intercal$	$(0.5,0.5)^\intercal$	$p=0, \phi=\infty$
3	$(0,0,-0.5,0,0.5,0.5,0,0,0.5,0.5)^\intercal$	$(0.5,0.5)^\intercal$	$p=1/2, \phi=1.0$
4	$(0,0,-0.5,0,0.5,0.5,0,0,0.49,0.49)^\intercal$	$(0.5,0.5)^\intercal$	$p=0, \phi=1.0204$
5	$(0,0,-0.5,0,1.00,1.00,0.5,0.5,0.5,0.5)^\intercal$	$(1.0,0.0)^\intercal$	$p=1/4, \phi=1.4142$
6	$(0,0,-0.5,0,0.25,0.25,0.5,0.5,0.5,0.5)^\intercal$	$(0.1,0.1)^\intercal$	$p=0, \phi=0.3451$
A	$(0,0,-0.25,0,0.75,0.75,0.5,0.5,0.5,0.5)^\intercal$	$(0.1,0.1)^\intercal$	$p=0, \phi=1.035$
B	$(0,0,0,0,0.25,0.25,0,0,0.25,0.25)^\intercal$	$(0,0)^\intercal$	$p=1/2, \phi=1.00$
C	$(0,0,0,0,0.25,0.25,0,0,0.24,0.24)^\intercal$	$(0,0)^\intercal$	$p=1/2, \phi=1.00$

Table 13: Parameters indexing the example models.

standardized effect size of treatment 1 versus treatment 2, treatment 1 versus treatment 3, and treatment 2 versus treatment 3 are all exactly equal in our examples. We report this as ϕ in Table 4.2. Tables 14 through 25 detail our results.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.844*	0.863*	0.911*	0.917*	0.940	0.922*	0.934*	0.900*	0.906*
PPE	0.949	0.948	0.933*	0.929*	0.939	0.823*	0.923*	0.870*	0.860*
ACI	0.971	0.979	0.966	0.969	0.963	0.971	0.956	0.974	0.977

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.835*	0.880*	0.931*	0.935*	0.931*	0.935*	0.941	0.927*	0.935*
PPE	0.944	0.942	0.948	0.948	0.940	0.863*	0.941	0.929*	0.928*
ACI	0.966	0.981	0.976	0.978	0.963	0.976	0.963	0.979	0.981

Table 14: Monte Carlo estimates of coverage probability of confidence intervals for $\beta^*_{1,0,1}$ (intercept term) at the 95% nominal level. Generative models two stages and **three** actions at the second stage. Estimates are constructed using 1000 datasets of size 150, 300 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.473*	0.473*	0.517*	0.518*	0.568	0.540*	0.555*	0.512*	0.511*
PPE	0.433	0.433	0.501*	0.501*	0.559	0.518*	0.557*	0.482*	0.481*
ACI	0.742	0.742	0.664	0.664	0.645	0.671	0.627	0.694	0.695

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.327*	0.327*	0.359*	0.359*	0.395*	0.380*	0.387	0.358*	0.358*
PPE	0.296	0.296	0.347	0.347	0.389	0.370*	0.386	0.343*	0.342*
ACI	0.517	0.516	0.461	0.461	0.448	0.453	0.423	0.468	0.469

Table 15: Monte Carlo estimates of the mean width of confidence intervals for $\beta^*_{1,0,1}$ (intercept term) at the 95% nominal level. Generative models two stages and **three** actions at the second stage. Estimates are constructed using 1000 datasets of size 150, 300 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.950	0.951	0.949	0.949	0.928*	0.944	0.942	0.955	0.955
PPE	0.951	0.951	0.951	0.951	0.925*	0.930*	0.937	0.954	0.954
ACI	0.987	0.986	0.977	0.976	0.941	0.955	0.947	0.969	0.969

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.935*	0.936	0.946	0.946	0.944	0.946	0.959	0.945	0.946
PPE	0.933*	0.935*	0.944	0.945	0.946	0.941	0.957	0.945	0.945
ACI	0.972	0.973	0.958	0.959	0.959	0.955	0.960	0.956	0.956

Table 16: Monte Carlo estimates of coverage probability of confidence intervals for $\beta^*_{1,0,2}$ (main effect of history) at the 95% nominal level. Generative models two stages and **three** actions at the second stage. Estimates are constructed using 1000 datasets of size 150, 300 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.342	0.342	0.344	0.344	0.433*	0.379	0.390	0.336	0.336
PPE	0.340	0.340	0.343	0.343	0.428*	0.372*	0.386	0.336	0.336
ACI	0.410	0.410	0.382	0.382	0.469	0.398	0.399	0.362	0.362

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.236*	0.236	0.237	0.237	0.302	0.263	0.269	0.231	0.231
PPE	0.234*	0.234*	0.236	0.236	0.299	0.259	0.268	0.231	0.231
ACI	0.280	0.280	0.261	0.261	0.327	0.270	0.273	0.242	0.242

Table 17: Monte Carlo estimates of the mean width of confidence intervals for $\beta^*_{1,0,2}$ (main effect of history) at the 95% nominal level. Generative models two stages and **three** actions at the second stage. Estimates are constructed using 1000 datasets of size 150, 300 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.933*	0.938	0.915*	0.921*	0.931*	0.907*	0.940	0.885*	0.895*
PPE	0.931*	0.932*	0.927*	0.919*	0.932*	0.883*	0.918*	0.858*	0.856*
ACI	0.999	0.999	0.968	0.970	0.964	0.972	0.964	0.970	0.971

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.950	0.948	0.926*	0.939	0.939	0.925*	0.943	0.927*	0.938
PPE	0.956	0.955	0.952	0.953	0.946	0.887*	0.943	0.937	0.938
ACI	0.999	0.999	0.965	0.971	0.961	0.974	0.966	0.967	0.970

Table 18: Monte Carlo estimates of coverage probability of confidence intervals for $\beta^*_{1,1,1}$ (main effect of treatment) at the 95% nominal level. Generative models two stages and **three** actions at the second stage. Estimates are constructed using 1000 datasets of size 150, 300 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.446*	0.446	0.518*	0.518*	0.567*	0.518*	0.557	0.508*	0.507*
PPE	0.415*	0.415*	0.500*	0.500*	0.557*	0.486*	0.548*	0.467*	0.465*
ACI	0.716	0.716	0.663	0.663	0.643	0.643	0.625	0.673	0.673

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.306	0.306	0.358*	0.358	0.395	0.372*	0.387	0.357*	0.357
PPE	0.284	0.284	0.346	0.346	0.389	0.345*	0.386	0.337	0.337
ACI	0.497	0.497	0.461	0.461	0.448	0.436	0.423	0.462	0.462

Table 19: Monte Carlo estimates of the mean width of confidence intervals for $\beta^*_{1,1,1}$ (main effect of treatment) at the 95% nominal level. Generative models two stages and **three** actions at the second stage. Estimates are constructed using 1000 datasets of size 150, 300 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.947	0.948	0.953	0.953	0.949	0.944	0.956	0.946	0.946
PPE	0.951	0.950	0.954	0.955	0.945	0.936	0.953	0.944	0.944
ACI	0.978	0.978	0.969	0.970	0.962	0.960	0.961	0.962	0.961

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.953	0.954	0.950	0.951	0.939	0.950	0.956	0.960	0.960
PPE	0.955	0.954	0.952	0.951	0.932*	0.940	0.952	0.959	0.960
ACI	0.985	0.986	0.975	0.975	0.955	0.954	0.958	0.968	0.968

Table 20: Monte Carlo estimates of coverage probability of confidence intervals for $\beta^*_{1,1,2}$ (interaction between history and treatment) at the 95% nominal level. Generative models two stages and **three** actions at the second stage. Estimates are constructed using 1000 datasets of size 150, 300 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.342	0.342	0.344	0.344	0.406	0.379	0.390	0.336	0.336
PPE	0.339	0.339	0.343	0.343	0.402	0.372	0.386	0.335	0.335
ACI	0.410	0.410	0.382	0.382	0.444	0.398	0.399	0.361	0.362

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.235	0.235	0.236	0.236	0.283	0.263	0.269	0.231	0.231
PPE	0.233	0.233	0.235	0.235	0.279*	0.259	0.268	0.231	0.231
ACI	0.280	0.280	0.261	0.261	0.308	0.269	0.272	0.242	0.242

Table 21: Monte Carlo estimates of the mean width of confidence intervals for $\beta^*_{1,1,2}$ (interaction between history and treatment) at the 95% nominal level. Generative models two stages and **three** actions at the second stage. Estimates are constructed using 1000 datasets of size 150, 300 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.934*	0.942	0.933*	0.934*	0.934*	0.920*	0.934*	0.907*	0.910*
PPE	0.941	0.939	0.935*	0.935*	0.929*	0.896*	0.918*	0.879*	0.881*
ACI	0.994	0.994	0.967	0.970	0.950	0.966	0.954	0.961	0.964

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.951	0.952	0.932*	0.936	0.946	0.939	0.948	0.930*	0.939
PPE	0.960	0.963	0.949	0.951	0.941	0.911*	0.947	0.932*	0.933*
ACI	0.998	0.997	0.968	0.972	0.965	0.970	0.966	0.968	0.972

Table 22: Monte Carlo estimates of coverage probability of confidence intervals for the contrast $\beta^*_{1,1,1} + \beta^*_{1,1,2}$ (effect of action for history = 1) at the 95% nominal level. Generative models two stages and **three** actions at the second stage. Estimates are constructed using 1000 datasets of size 150, 300 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.561*	0.561	0.620*	0.620*	0.691*	0.639*	0.676*	0.607*	0.607*
PPE	0.536	0.536	0.605*	0.605*	0.686*	0.612*	0.669*	0.575*	0.574*
ACI	0.833	0.833	0.767	0.767	0.734	0.762	0.741	0.773	0.774

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.386	0.386	0.428*	0.428	0.481	0.454	0.469	0.426*	0.425
PPE	0.367	0.368	0.418	0.418	0.478	0.430*	0.468	0.409*	0.409*
ACI	0.577	0.577	0.532	0.532	0.510	0.517	0.503	0.530	0.530

Table 23: Monte Carlo estimates of the mean width of confidence intervals for the contrast $\beta^*_{1,1,1} + \beta^*_{1,1,2}$ (effect of action for history = 1) at the 95% nominal level. Generative models two stages and **three** actions at the second stage. Estimates are constructed using 1000 datasets of size 150, 300 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.947	0.948	0.937	0.939	0.931*	0.929*	0.944	0.914*	0.914*
PPE	0.948	0.949	0.940	0.939	0.929*	0.911*	0.928*	0.893*	0.896*
ACI	0.997	0.997	0.977	0.979	0.967	0.977	0.961	0.978	0.979

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.944	0.942	0.923*	0.934*	0.938	0.936	0.942	0.931*	0.935*
PPE	0.948	0.948	0.948	0.952	0.947	0.910*	0.937	0.937	0.936
ACI	0.996	0.996	0.964	0.970	0.960	0.961	0.960	0.967	0.969

Table 24: Monte Carlo estimates of coverage probability of confidence intervals for the contrast $\beta^*_{1,1,1} - \beta^*_{1,1,2}$ (effect of action for history = -1) at the 95% nominal level. Generative models two stages and **three** actions at the second stage. Estimates are constructed using 1000 datasets of size 150, 300 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.561	0.561	0.624	0.623	0.702*	0.644*	0.682	0.610*	0.610*
PPE	0.535	0.535	0.607	0.607	0.686*	0.610*	0.672*	0.576*	0.575*
ACI	0.834	0.833	0.767	0.767	0.816	0.769	0.751	0.777	0.777

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.386	0.386	0.430*	0.430*	0.490	0.458	0.473	0.425*	0.425*
PPE	0.367	0.367	0.418	0.419	0.479	0.432*	0.471	0.409	0.408
ACI	0.578	0.578	0.531	0.531	0.569	0.522	0.510	0.530	0.530

Table 25: Monte Carlo estimates of the mean width of confidence intervals for the contrast $\beta^*_{1,1,1} - \beta^*_{1,1,2}$ (effect of action for history = -1) at the 95% nominal level. Generative models two stages and **three** actions at the second stage. Estimates are constructed using 1000 datasets of size 150, 300 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

4.3 Models with three stages

Here, we present results using another suite of examples, again similar to those of Chakraborty et al. (2009), but that have three stages of treatment, with binary treatments at each stage. These models are defined as follows:

- $X_i \in \{-1, 1\}$, $A_i \in \{-1, 1\}$ for $i \in \{1, 2, 3\}$
- $P(A_i = 1) = P(A_i = -1) = 0.5$ for $i \in \{1, 2, 3\}$
- $P(X_1 = 1) = P(X_1 = -1) = 1/2$,
  $P(X_{i+1} = 1 | X_i, A_i) = \text{expit}(\delta_1 X_i + \delta_2 A_i)$ for $i \in \{1, 2\}$
- $Y_1 \triangleq Y_2 \triangleq 0$
  $$\begin{aligned} Y_3 = \xi_1 + \xi_2 X_1 + \xi_3 A_1 + \xi_4 X_1 A_1 + \\ \xi_5 A_2 + \xi_6 X_2 A_2 + \xi_7 A_1 A_2 + \\ \xi_8 A_3 + \xi_9 X_3 A_3 + \xi_{10} A_2 A_3 + \epsilon \end{aligned}$$
  $\epsilon \sim N(0, 1)$

where $\text{expit}(x) = \text{e}^x/(1+\text{e}^x)$. This class is parameterized by twelve values $\xi_1, \xi_2, ..., \xi_{10}, \delta_1, \delta_2$. The analysis model uses histories defined by:

$$H_{3,0} = (1, X_1, A_1, X_1A_1, X_2, A_2, X_2A_2, A_1A_2, X_3)^\intercal \tag{36}$$

$$H_{3,1} = (1, X_3, A_2)^\intercal \tag{37}$$

$$H_{2,0} = (1, X_1, A_1, X_1A_1, X_2)^\intercal \tag{38}$$

$$H_{2,1} = (1, X_2, A_1)^\intercal \tag{39}$$

$$H_{1,0} = (1, X_1)^\intercal \tag{40}$$

$$H_{1,1} = (1, X_1)^\intercal. \tag{41}$$

Example	ξ	δ	Stage 2&3 Regularity
1	$(0,0,0,0,0,0,0,0,0,0)^\intercal$	$(0.5,0.5)^\intercal$	$p=1, \phi=0/0$
2	$(0,0,0,0,0.01,0,0,0.01,0,0)^\intercal$	$(0.5,0.5)^\intercal$	$p=0, \phi=\infty$
3	$(0,0,-0.5,0,0,0,0.5,0.5,0,0.5)^\intercal$	$(0.5,0.5)^\intercal$	$p=1/2, \phi=1.003$
4	$(0,0,-0.5,0,0,0,0.49,0.5,0,0.49)^\intercal$	$(0.5,0.5)^\intercal$	$p=0, \phi=1.014$
5	$(0,0,-0.5,0,0.5,0.5,0.5,1.0,0.5,0.5)^\intercal$	$(1.0,0.0)^\intercal$	$p=1/4, \phi=1.40$
6	$(0,0,-0.5,0,0.12,0.48,0.50,0.25,0.5,0.5)^\intercal$	$(0.1,0.1)^\intercal$	$p=0, \phi=0.349$
A	$(0,0,-0.25,0,0.36,0.49,0.50,0.75,0.5,0.5)^\intercal$	$(0.1,0.1)^\intercal$	$p=0, \phi=1.05$
B	$(0,0,0,0,0,0,0.25,0.25,0,0.25)^\intercal$	$(0,0)^\intercal$	$p=1/2, \phi=1.00$
C	$(0,0,0,0,0,0,0.24,0.25,0,0.24)^\intercal$	$(0,0)^\intercal$	$p=0, \phi=1.03$

Table 26: Parameters indexing the example models.

The values of the constants $\xi_1, ..., \xi_{10}$ and δ_1, δ_2 in Examples 1 through 7 given in Table 26. Since the third stage of these models has the same structure and parameters as the second stage of the two-stage models in Chakraborty et al. (2009), the measures p and ϕ of non-regularity for the final stages in both suites of examples are exactly the same.

We have chosen parameters ξ_5, ξ_6, ξ_7 so that the measures of non-regularity at stage 2 are exactly the same as they are at stage 3. The coefficients and regularity properties for both stages are given in Table 26.

Tables 27 through 38 detail our results for these models. Note that the original work of Chakraborty et al. (2009) did not define the Soft Threshold (ST) method for more than two stages, so we make the most obvious extension. To produce a ST confidence interval for $c^\intercal\beta_1^*$ for a 3-stage problem, we bootstrap a shrunken estimate of β_1^* that is based on the standard Q-learning estimate of β_2^*. We then use the hybrid bootstrap to produce the CI.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.836*	0.868*	0.918*	0.920*	0.926*	0.921*	0.935*	0.908*	0.908*
PPE	0.924*	0.937	0.932*	0.931*	0.929*	0.892*	0.935*	0.887*	0.884*
ST	0.938	0.946	0.912*	0.907*	0.886*	0.558*	0.762*	0.710*	0.692*
ACI	0.950	0.962	0.964	0.966	0.944	0.963	0.958	0.962	0.963

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.833*	0.879*	0.933*	0.937	0.951	0.934*	0.944	0.934*	0.934*
PPE	0.931*	0.949	0.952	0.950	0.955	0.904*	0.952	0.926*	0.925*
ST	0.930*	0.945	0.944	0.936	0.910*	0.615*	0.810*	0.871*	0.856*
ACI	0.952	0.970	0.971	0.971	0.971	0.957	0.959	0.971	0.972

Table 27: Monte Carlo estimates of coverage probability of confidence intervals for $\beta^*_{1,0,1}$ (intercept term) at the 95% nominal level. Generative models have three stages and two actions per stage. Estimates are constructed using 1000 datasets of size 150, 300 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.503*	0.503*	0.563*	0.563*	0.660*	0.591*	0.625*	0.555*	0.555*
PPE	0.481*	0.481	0.560*	0.560*	0.660*	0.572*	0.641*	0.531*	0.529*
ST	0.438	0.438	0.588*	0.590*	0.699*	0.603*	0.682*	0.572*	0.570*
ACI	0.657	0.657	0.659	0.659	0.728	0.679	0.694	0.673	0.674

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.345*	0.346*	0.389*	0.390	0.459	0.417*	0.428	0.388*	0.389*
PPE	0.330*	0.330	0.387	0.387	0.459	0.401*	0.446	0.381*	0.380*
ST	0.298*	0.298	0.397	0.398	0.472*	0.430*	0.447*	0.415*	0.416*
ACI	0.453	0.453	0.456	0.456	0.508	0.459	0.467	0.459	0.460

Table 28: Monte Carlo estimates of the mean width of confidence intervals for $\beta^*_{1,0,1}$ (intercept term) at the 95% nominal level. Generative models have three stages and two actions per stage. Estimates are constructed using 1000 datasets of size 150, 300 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.952	0.952	0.951	0.951	0.942	0.955	0.969	0.954	0.954
PPE	0.956	0.955	0.948	0.949	0.941	0.951	0.966	0.957	0.957
ST	0.952	0.952	0.956	0.956	0.945	0.955	0.965	0.959	0.957
ACI	0.968	0.968	0.962	0.962	0.957	0.962	0.971	0.968	0.967

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.963	0.964	0.969	0.968	0.956	0.953	0.957	0.964	0.965
PPE	0.964	0.963	0.969	0.969	0.956	0.949	0.956	0.965	0.966
ST	0.967	0.965	0.968	0.968	0.963	0.945	0.959	0.963	0.963
ACI	0.975	0.973	0.974	0.974	0.964	0.955	0.958	0.971	0.971

Table 29: Monte Carlo estimates of coverage probability of confidence intervals for $\beta^*_{1,0,2}$ (main effect of history) at the 95% nominal level. Generative models have three stages and two actions per stage. Estimates are constructed using 1000 datasets of size 150, 300 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.339	0.340	0.340	0.340	0.441	0.384	0.388	0.336	0.336
PPE	0.338	0.338	0.340	0.340	0.438	0.379	0.385	0.336	0.336
ST	0.337	0.337	0.340	0.340	0.451	0.391	0.397	0.336	0.336
ACI	0.367	0.367	0.355	0.355	0.465	0.396	0.394	0.352	0.352

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.234	0.234	0.233	0.234	0.307	0.265	0.267	0.231	0.231
PPE	0.233	0.233	0.234	0.234	0.306	0.263	0.266	0.231	0.231
ST	0.233	0.233	0.234	0.234	0.312	0.271	0.271	0.231	0.231
ACI	0.249	0.249	0.242	0.242	0.324	0.269	0.269	0.237	0.237

Table 30: Monte Carlo estimates of the mean width of confidence intervals for $\beta^*_{1,0,2}$ (main effect of history) at the 95% nominal level. Generative models have three stages and two actions per stage. Estimates are constructed using 1000 datasets of size 150, 300 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.949	0.952	0.926*	0.926*	0.945	0.914*	0.944	0.897*	0.896*
PPE	0.952	0.955	0.933*	0.932*	0.942	0.911*	0.923*	0.850*	0.853*
ST	0.945	0.946	0.913*	0.913*	0.954	0.924*	0.927*	0.754*	0.755*
ACI	0.981	0.983	0.949	0.948	0.962	0.958	0.957	0.933*	0.934*

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.957	0.953	0.929*	0.936	0.935*	0.913*	0.949	0.928*	0.924*
PPE	0.960	0.962	0.934*	0.940	0.945	0.903*	0.931*	0.903*	0.898*
ST	0.953	0.951	0.920*	0.923*	0.947	0.911*	0.906*	0.872*	0.881*
ACI	0.981	0.979	0.952	0.956	0.958	0.943	0.954	0.946	0.947

Table 31: Monte Carlo estimates of coverage probability of confidence intervals for $\beta^*_{1,1,1}$ (main effect of treatment) at the 95% nominal level. Generative models have three stages and two actions per stage. Estimates are constructed using 1000 datasets of size 150, 300 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.406	0.406	0.485*	0.485*	0.530	0.496*	0.519	0.474*	0.472*
PPE	0.386	0.386	0.466*	0.465*	0.523	0.477*	0.511*	0.427*	0.425*
ST	0.400	0.400	0.489*	0.489*	0.544	0.517*	0.545*	0.484*	0.481*
ACI	0.476	0.476	0.530	0.530	0.574	0.553	0.558	0.525*	0.524*

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.279	0.279	0.335*	0.334	0.371*	0.352*	0.362	0.333*	0.332*
PPE	0.265	0.265	0.321*	0.321	0.366	0.338*	0.360*	0.311*	0.309*
ST	0.274	0.275	0.329*	0.328*	0.375	0.368*	0.373*	0.344*	0.343*
ACI	0.323	0.323	0.364	0.363	0.401	0.381	0.380	0.362	0.362

Table 32: Monte Carlo estimates of the mean width of confidence intervals for $\beta^*_{1,1,1}$ (main effect of treatment) at the 95% nominal level. Generative models have three stages and two actions per stage. Estimates are constructed using 1000 datasets of size 150, 300 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.942	0.941	0.947	0.945	0.943	0.939	0.953	0.948	0.948
PPE	0.948	0.946	0.946	0.947	0.945	0.934*	0.951	0.949	0.949
ST	0.948	0.948	0.945	0.945	0.945	0.942	0.949	0.942	0.942
ACI	0.959	0.959	0.954	0.954	0.954	0.945	0.955	0.959	0.959

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.948	0.949	0.952	0.953	0.954	0.947	0.961	0.953	0.954
PPE	0.951	0.951	0.949	0.949	0.954	0.946	0.959	0.952	0.953
ST	0.950	0.951	0.948	0.950	0.956	0.945	0.963	0.951	0.951
ACI	0.967	0.966	0.959	0.959	0.964	0.949	0.962	0.960	0.960

Table 33: Monte Carlo estimates of coverage probability of confidence intervals for $\beta^*_{1,1,2}$ (interaction between history and treatment) at the 95% nominal level. Generative models have three stages and two actions per stage. Estimates are constructed using 1000 datasets of size 150, 300 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.339	0.339	0.342	0.342	0.426	0.385	0.389	0.337	0.337
PPE	0.338	0.338	0.341	0.341	0.421	0.379*	0.386	0.337	0.337
ST	0.338	0.338	0.343	0.343	0.434	0.392	0.397	0.337	0.337
ACI	0.365	0.365	0.357	0.357	0.447	0.396	0.395	0.352	0.353

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.232	0.232	0.234	0.234	0.296	0.264	0.266	0.230	0.230
PPE	0.231	0.231	0.233	0.233	0.293	0.262	0.265	0.230	0.230
ST	0.232	0.232	0.234	0.234	0.299	0.270	0.270	0.231	0.231
ACI	0.247	0.247	0.242	0.242	0.311	0.268	0.268	0.237	0.237

Table 34: Monte Carlo estimates of the mean width of confidence intervals for $\beta^*_{1,1,2}$ (interaction between history and treatment) at the 95% nominal level. Generative models have three stages and two actions per stage. Estimates are constructed using 1000 datasets of size 150, 300 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.942	0.940	0.935*	0.931*	0.946	0.930*	0.930*	0.913*	0.913*
PPE	0.951	0.949	0.935*	0.931*	0.945	0.921*	0.919*	0.877*	0.877*
ST	0.947	0.948	0.916*	0.922*	0.947	0.928*	0.925*	0.814*	0.819*
ACI	0.975	0.976	0.949	0.949	0.953	0.957	0.946	0.939	0.944

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.953	0.956	0.940	0.947	0.946	0.942	0.946	0.933*	0.934*
PPE	0.959	0.960	0.946	0.950	0.945	0.933*	0.944	0.920*	0.922*
ST	0.954	0.957	0.930*	0.931*	0.949	0.939	0.935*	0.896*	0.899*
ACI	0.979	0.980	0.960	0.964	0.955	0.960	0.959	0.954	0.957

Table 35: Monte Carlo estimates of coverage probability of confidence intervals for the contrast $\beta^*_{1,1,1} + \beta^*_{1,1,2}$ (effect of action for history = 1) at the 95% nominal level. Generative models have three stages and two actions per stage. Estimates are constructed using 1000 datasets of size 150, 300 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.529	0.529	0.593*	0.592*	0.665	0.624*	0.644*	0.581*	0.579*
PPE	0.511	0.511	0.575*	0.575*	0.664	0.607*	0.639*	0.542*	0.541*
ST	0.522	0.523	0.595*	0.595*	0.683	0.645*	0.669*	0.589*	0.586*
ACI	0.605	0.605	0.641	0.641	0.693	0.682	0.682	0.638	0.637

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.363	0.363	0.407	0.407	0.465	0.438	0.446	0.405*	0.404*
PPE	0.352	0.352	0.396	0.396	0.464	0.426*	0.445	0.386*	0.385*
ST	0.360	0.360	0.403*	0.402*	0.472	0.454	0.458*	0.415*	0.414*
ACI	0.411	0.411	0.438	0.438	0.483	0.467	0.463	0.436	0.436

Table 36: Monte Carlo estimates of the mean width of confidence intervals for the contrast $\beta^*_{1,1,1} + \beta^*_{1,1,2}$ (effect of action for history = 1) at the 95% nominal level. Generative models have three stages and two actions per stage. Estimates are constructed using 1000 datasets of size 150, 300 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.940	0.939	0.929*	0.926*	0.942	0.935*	0.948	0.918*	0.917*
PPE	0.936	0.937	0.929*	0.929*	0.940	0.915*	0.927*	0.888*	0.888*
ST	0.934*	0.934*	0.920*	0.922*	0.955	0.929*	0.934*	0.815*	0.817*
ACI	0.965	0.966	0.947	0.951	0.961	0.952	0.961	0.940	0.939

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.943	0.945	0.932*	0.936	0.935*	0.929*	0.941	0.931*	0.934*
PPE	0.946	0.945	0.937	0.939	0.943	0.908*	0.926*	0.922*	0.920*
ST	0.945	0.947	0.930*	0.930*	0.955	0.919*	0.921*	0.895*	0.896*
ACI	0.973	0.970	0.949	0.953	0.956	0.947	0.952	0.949	0.950

Table 37: Monte Carlo estimates of coverage probability of confidence intervals for the contrast $\beta^*_{1,1,1} - \beta^*_{1,1,2}$ (effect of action for history = -1) at the 95% nominal level. Generative models have three stages and two actions per stage. Estimates are constructed using 1000 datasets of size 150, 300 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

$N = 150$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.530	0.530	0.595*	0.594*	0.693	0.631*	0.652	0.582*	0.580*
PPE	0.513	0.513	0.577*	0.577*	0.679	0.609*	0.641*	0.544*	0.543*
ST	0.524*	0.524*	0.600*	0.599*	0.708	0.653*	0.678*	0.590*	0.587*
ACI	0.605	0.605	0.643	0.642	0.755	0.687	0.690	0.639	0.638

$N = 300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
CPB	0.363	0.363	0.409*	0.408	0.483*	0.443*	0.451	0.405*	0.404*
PPE	0.351	0.351	0.398	0.398	0.474	0.427*	0.448*	0.387*	0.385*
ST	0.359	0.359	0.405*	0.404*	0.486	0.459*	0.463*	0.414*	0.414*
ACI	0.410	0.410	0.440	0.440	0.527	0.470	0.469	0.436	0.436

Table 38: Monte Carlo estimates of the mean width of confidence intervals for the contrast $\beta^*_{1,1,1} - \beta^*_{1,1,2}$ (effect of action for history = -1) at the 95% nominal level. Generative models have three stages and two actions per stage. Estimates are constructed using 1000 datasets of size 150, 300 are drawn from each model, and 1000 bootstraps drawn from each dataset. Estimates significantly below 0.95 at the 0.05 level are marked with $*$. Models are designated NR = non-regular, NNR = near-non-regular, R = regular.

References

P.J. Bickel and D.A. Freedman. Some asymptotic theory for the bootstrap. *The Annals of Statistics*, pages 1196–1217, 1981.

B. Chakraborty, S. Murphy, and V. Strecher. Inference for non-regular parameters in optimal dynamic treatment regimes. *Statistical Methods in Medical Research*, 19(3), 2009.

Jeesen Chen. Notes on the bias-variance trade-off phenomenon. *A Festschrift for Herman Rubin: Institute of Mathematical Statistics*, 45:207–217, 2004.

Sándor Csörgő and Andrew Rosalsky. A survey of limit laws for bootstrapped sums. *International Journal of Mathematics and Mathematical Statistics*, 45:2835–2861, 2003.

Hani Doss and Jayaram Sethuraman. The price of bias reduction when there is no unbiased estimate. *Annals of Statistics*, 17(1):440–442, 1989.

Keisuke Hirano and Jack Porter. Impossibility results for nondifferentiable functionals. Mpra paper, University Library of Munich, Germany, 2009. URL `http://econpapers.repec.org/RePEc:pra:mprapa:15990`.

Michael R. Kosorok. *Introduction to empirical processes and semiparametric inference*. Springer, 2008.

Richard C. Liu and Lawrence D. Brown. Nonexistence of informative unbiased estimators in singular problems. *Annals of Statistics*, 21(1):1–13, 1993.

E.E.M. Moodie, T.S. Richardson, and D.A. Stephens. Estimating optimal dynamic regimes: Correcting bias under the null. *Biometrics*, 63(2):447–455, 2010.

J.M. Robins. Optimal structural nested models for optimal sequential decisions. In *Pro-*

ceedings of the Second Seattle Symposium in Biostatistics: Analysis of Correlated Data, 2004.